Herausgeber

F. PUENTE LEÓN
M. HEIZMANN

FORUM BILDVERARBEITUNG 2014

KIT Scientific Publishing

F. Puente León | M. Heizmann (Hrsg.)

FORUM BILDVERARBEITUNG 2014

FORUM BILDVERARBEITUNG 2014

F. Puente León
M. Heizmann
(Hrsg.)

KIT Scientific Publishing

Impressum

KIT Scientific Publishing

Karlsruher Institut für Technologie (KIT)
KIT Scientific Publishing
Straße am Forum 2
D-76131 Karlsruhe

www.ksp.kit.edu

Print on Demand 2014

ISBN 978-3-7315-0284-5
DOI 10.5445/KSP/1000043608

Vorwort

Bildverarbeitung spielt in vielen Bereichen der Technik eine Schlüsselrolle. Etwa in der Qualitätssicherung industrieller Prozesse oder bei der Fahrerassistenz haben sich Bildverarbeitungssysteme einen unverzichtbaren Platz erobert. Dennoch werden in der Bildverarbeitung weiterhin erhebliche Fortschritte erzielt: Sie werden auf der Seite der Hardware durch Weiterentwicklungen im Bereich der Sensortechnik, der Datenübertragung und durch die Zunahme der Leistungsfähigkeit von Rechnersystemen getragen. Auf der Seite der Signal- und Informationsverarbeitung sind leistungsfähige mathematische Verfahren und effiziente Algorithmen zur Verarbeitung der von Kameras erfassten Bildsignale wichtige Schwerpunkte aktueller Forschungs- und Entwicklungsarbeit.

Das „Forum Bildverarbeitung“ hat sich zum Ziel gesetzt, aktuelle Trends auf den Gebieten der Bildgewinnung, -verarbeitung und -auswertung aufzugreifen und zum fachlichen Austausch zwischen den Teilnehmern beizutragen. Die Beiträge wurden vom Programmausschuss ausgewählt, welcher vom Fachausschuss 3.51 „Bildverarbeitung in der Mess- und Automatisierungstechnik“ der VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik (GMA) berufen wurde. Sie umfassen die folgenden Schwerpunkte:

- Auslegung von Bildverarbeitungssystemen,
- Bildgewinnung,
- Computergrafik für die automatische Sichtprüfung,
- Mathematische Modelle und Verfahren,
- Medizin und Biologie,
- Oberflächenprüfung,
- Positionsbestimmung sowie
- 3D-Messung mit strukturierter Beleuchtung.

Das Forum Bildverarbeitung möchte einen Beitrag zur Weiterentwicklung dieser wichtigen und zukunftsträchtigen Disziplin leisten. Es richtet sich an Fachleute, die sich in der industriellen Entwicklung, in

der Forschung oder der Lehre mit Bildverarbeitungssystemen befassen, und bietet eine Plattform für den Wissens- und Erfahrungsaustausch zwischen Wissenschaftlern und Anwendern.

November 2014 F. Puente León und M. Heizmann

Wissenschaftliche Leitung

Prof. Dr.-Ing. M. Heizmann	Hochschule Karlsruhe, Fraunhofer IOSB
Prof. Dr.-Ing. F. Puente León	Karlsruher Institut für Technologie

Programmausschuss

Prof. Dr. C. Bach	NTB, CH-Buchs
Prof. Dr.-Ing. J. Beyerer	Fraunhofer IOSB Karlsruhe
Prof. Dr. K. Donner	Universität Passau
Dr. rer. nat. J. Eggert	Honda Research Institute, Offenbach
Prof. Dr. A. Heinrich	Hochschule Aalen
Prof. Dr.-Ing. M. Heizmann	Fraunhofer IOSB Karlsruhe
Prof. Dr. B. Jähne	Universität Heidelberg
Prof. Dr.-Ing. T. Längle	Fraunhofer IOSB Karlsruhe
Dr.-Ing. E. Marquardt	VDI e.V., Düsseldorf
Dipl.-Ing. M. Maurer	Vitronic Dr.-Ing. Stein GmbH
Prof. Dr. R. Neubecker	Hochschule Darmstadt
Prof. Dr. W. Osten	Universität Stuttgart
Dr.-Ing. C. Otto	Daimler AG, Stuttgart
Prof. Dr.-Ing. F. Puente León	Karlsruher Institut für Technologie
Prof. Dr. F. Salazar	Universidad Politécnica de Madrid
Prof. Dr.-Ing. R. Schmitt	RWTH Aachen
Dipl.-Ing. M. Stelzl	Schott AG, Mainz
Prof. Dr.-Ing. C. Stiller	Karlsruher Institut für Technologie
Prof. Dr.-Ing. R. Tutsch	Technische Universität Braunschweig
Dr.-Ing. S. Werling	Fraunhofer IOSB Karlsruhe
Dipl.-Ing. S. Wienand	ISRA VISION AG, Darmstadt
Dr.-Ing. V. Willert	TU Darmstadt

Inhaltsverzeichnis

Vorwort v

Auslegung von Bildverarbeitungssystemen

Capability of classifying inspection systems 1
R. Neubecker

Struktur zur Auswahl und Implementierung von bildverarbeitenden Systemen in der Fertigungsautomation 15
A. Grote und E. Schwab

Self-optimized adaptive algorithm solutions for vision systems ... 27
T. Rashba and S. Richter

Computergrafik für die automatische Sichtprüfung

Realization and evaluation of image processing tasks based on synthetic sensor data: 2 use cases 35
S. Irgenfried, F. Dittrich and H. Wörn

Synthetic image acquisition and procedural modeling for automated optical inspection (AOI) systems 47
M. Retzlaff, J. Stabenow and C. Dachsbacher

Mathematische Modelle und Verfahren

Verdeckungs- und affin-invarianter Regionendetektor basierend auf Farb- und Frequenzinformation 61
P. Hernández Mesa, R. Heiman und F. Puente León

Fast image super-resolution utilizing convolutional neural networks 73
H. Soyer and C. Osendorfer

Optimierung der *Fast Radial Symmetry Detection* für eine echtzeitfähige Kreisdetektion 85
S. Eickeler und M. Valdenegro

Exploitation of GPS control points in low-contrast IR imagery for homography estimation 97
P. Dunau

Parameter-learning for color sorting of bulk materials using genetic algorithms 107
M. Richter und J. Beyerer

Mehrdimensionale Merkmale zur Augendetektion 119
S. Vater und F. Puente León

Oberflächenprüfung

Segmentierung unterschiedlich stark ausgeprägter Welligkeiten auf lackierten Oberflächen 129
M. Vogelbacher, M. Ziebarth, S. Olawsky und J. Beyerer

Digitalisierung dreidimensionaler CFK-Halbzeuge zur Fehlstellenklassifizierung am Beispiel der Faserwelligkeit 141
P. Kosse, T. Fürtjes und R. Schmitt

Sichtbarkeit von Dellen und Beulen auf spiegelnden Oberflächen . 153
M. Ziebarth, M. Heizmann und J. Beyerer

Medizin und Biologie

Bildverarbeitungsbasierte Quantifizierung der Konfluenz von Stammzellkolonien zur Prozesssteuerung in einer Bioproduktionsanlage 167
F. Schenk, C. Kowalski und R. Schmitt

Automatisierte Beurteilung der Schädigungssituation bei Patienten mit altersbedingter Makuladegeneration (AMD) 179
S. Kahl, M. Ritter und P. Rosenthal

Superpixel-gestützte Klassifikation von Stechmückengattungen mit der Bags-of-Features-Methode 191
P. Grigoriev, J. Jäger, C. Kornek, V. Wolff und K. Fricke-Neuderth

Zweistufige Anwendung der Saliency-Methodik zur Stechmückendetektion 203
J. Jäger, V. Wolff und K. Fricke-Neuderth

Bildgewinnung

Industrielle Sortierung von Mineralen anhand von hyperspektralen Fluoreszenzaufnahmen – Potenzialbewertung . . . 215
S. Bauer und F. Puente León

Ellipsometrie an gekrümmten Oberflächen 227
C. Negara und M. Hartrumpf

Verbesserung von Positionsbestimmungen mittels holografischer Mehrpunktgenerierung 239
T. Haist, M. Gronle, T. Arnold, D. Bui und W. Osten

Positionsbestimmung

Positionierungsverifikation komplexer Großbauteile in der Roboterzelle zur Erweiterung eines Prozessführungssystems für die Automatisierung von MRO-Prozessen 249
J. P. Steinbach, T. Ernst, A. Fay und F. Hartung

Kamerabasiertes Referenzsystem für Fahrerassistenzsysteme 261
T. Kubertschak, M. Wittenzellner und M. Maehlisch

Richtungsabhängige Personendetektion und -verfolgung 273
J. Pallauf und F. Puente León

3D-Messung mit strukturierter Beleuchtung

Streifenprojektionsgenauigkeit mit Kinect-Rate – 3D-Sensorik für schnelle, dichte und genaue Formvermessung 285
M. Schaffer und M. Große

A different approach to multi-period phase shift 295
T. Dunker and S. Luther

Capability of classifying inspection systems

Ralph Neubecker

Hochschule Darmstadt, Optotechnik und Bildverarbeitung, FB MN,
Schöfferstr. 3, 64295 Darmstadt

Abstract Today, image processing systems play an important role in industrial production, one application being inspection for quality control. Still, for classifying systems, there are no widely accepted standards to quantify their capability. A scheme for this purpose will be presented, covering the whole inspection process. It allows to derive necessary performance figures for the image processing system. Aspects of an empirical determination of those figures, as necessary for a validation, are also outlined.

1 Introduction

Image processing systems are well-established elements in modern manufacturing. In particular, quality control in high throughput mass production would not have reached the present level, if it would still rely on visual inspection by human beings. Only automatic systems allow reliable and high-speed 100%-inspection. One particular function of such systems is classification, as e.g. used in surface inspection. In spite of the practical importance of such systems, there is no standard to quantitatively determine their capability for a specific inspection task.

For measuring systems, one can rely on on established standards. The quality of a such a system is described by a measurement uncertainty and its capability for a specific task is judged by capability indexes or similar parameters [1–5]. For classifying systems instead, which give attributive results instead of continuous measurands, there are no similar definitions nor practices.

Some procedures for attributive measurements are described in [4,5]. However, aiming at human inspectors, those procedures cannot really be applied to automatic inspection systems. Moreover, the classification performance is summarized in a single figure of merit (Cohen's Kappa).

There are a number of comparable single measures, like Pearsons Contingency Coefficient, or Cramers V [6–8]. Therefore, the choice for Cohen's Kappa is somewhat arbitrary, also the choice for its limit values. But the main point is that condensing the classification performance to a single number does not allow to model a whole inspection process.

In this paper, an approach will be presented, focusing on inspection systems [9,10]. It is based on a model of the complete inspection process, arriving at the false accept rate and false reject rate, which describe the accuracy of the final inspection result. As will be shown, these rates depend not only on the core performance of the inspection system, but also on factors under the responsibility of the user. This aspect is of practical importance, as a common understanding of the corresponding relations may ease acceptance and validation procedures.

In the remainder of this section, the inspection process will be regarded in detail. The next section is devoted to the theoretical modeling. Before concluding, application aspects are covered, particularly the use for scenario calculations and questions concerning empirical validation.

The basis of the present paper is the use of a classifying image processing system for quality control purposes. Typical applications of this kind are final inspection systems, looking for product defects. We will in particular consider systems inspecting separate products (*parts*), in contrast to a continuous production like paper, fabrics etc.. As already indicated, the whole inspection process is not only determined by the inspection system itself. Other significant influence factors are the quality criteria, by which is decided what makes a part under test to be scrap. Also, the production process has strong impact in the sense that it determines the frequency and the kind of defects reaching the inspection system. A crucial point is that a part under test may carry many potential defects. The more it carries, the better the inspection system must be to yield a correct inspection result.

The inspection system, as it is understood here, operates based on events: it has to find any potential defect on the part under test and then needs to determine, what kind of defect this has been. In many (but not in all) cases, it is not clear, where potential defects are located on the part. Hence, the system has to identify critical spots, which may be a defect - this step is called *detection* in the following. The corresponding image regions are then *classified* into predefined classes. It may however turn out, that the detected region represents something harmless, like a

dust particle. Hence, anything that should or could be detected as potential defect is summarized under the notion *event*: real product defects and any other local feature with similar optical properties. What will be counted as event also depends on the inspection hardware design, in particular on the illumination concept. In any case, all detected events will be classified, hence, when setting up the classifier, all possible event types (classes) must be named.

After the main classification of an event, it still has to be decided if this event is relevant for the product quality. This second classification step will be called *qualification* in the following. For simplicity we will assume that there are only two quality classes, namely *ok* and *nok* = not ok (in this notation, real defects are nok-events).

In general, the final decision on the quality of the part under test depends on the number and class of the events found. In the simplest case there are only two quality classes for parts: OK and NOK (capital letters are used to distinguish between event quality and part quality). The simplest possible quality rule is that a part will be NOK if it carries one or more defects. A good part (OK) may carry many ok-events, like dust particles or scratches, which are detectable but are not quality-relevant.

The user of an inspection system is only interested in the accuracy of the final inspection outcome. In our case with only two quality classes, there are four cases: correct inspection of good parts ($OK \rightarrow OK$), correct inspection of bad parts ($NOK \rightarrow NOK$), false reject ($OK \rightarrow NOK$) and false accept ($NOK \rightarrow OK$). We describe this by a *inspection rate* Q_{LM}, where the indices L and M stand for the two part quality classes.

2 Inspection process model

2.1 Single events

As pointed out, the main functions of a inspection processing system are detection and classification. Consequently, its core performance can be defined by a *detection rate* d_i and by a classification rate c_{ij}, describing the relative ratio of correctly detected and correctly classified events. The indexes i, j refer to the type (class) of the event. The classification rate also describes cross-classification, assigning the event from real class i to class j.

Both numbers can be combined to a *recognition rate* $r_{ij} = d_i \cdot c_{ij}$, completely describing the inspection system. In some cases, detection is not necessary, since the location of potential defects is known in advance, e.g. in the case of a control of correct filling of a blister packaging. In our context, this means $d_i = 1$.

The detection- and classification rates only tell us, which percentage of the events are correctly detected and classified - but the absolute number of undetected or mis-classified events still depends on the frequency by which the events occur. We describe this by the average number h_i of each particular event of type i on a single part under test.

Finally, we can define a *qualification rate* q_{lm}, describing the correct assignment of the event quality class, where the indexes l and m stand for ok and nok.

In order to relate those rates to the actual production situation, one can look at *weighted rates*, e.g. $r^{\star}_{ij} = h_i \cdot r_{ij}$ and $q^{\star}_{lm} = h_l \cdot c_{lm}$, which include the frequency of occurrence of the regarded events.

2.2 Parts

Events per part

So far, only the average number of defects per part has been stated. The actual number on the part under test may follow an unknown statistics. This distribution may either be determined empirically or taken from an adequate model. Under the condition of independent events - which excludes correlations caused by the production process, e.g. the occurrence of defect nests - one may assume a Poisson distribution

$$P_P(k) = \frac{h^k}{k!} e^{-h} \tag{1.1}$$

for the probability that k events are found on a part, when h is the average number.

Inspecting a good part

A good part contains only ok-events. Either all of them are *qualified* correctly, or one or more of them are mis-qualified to be of nok-type (false reject).

Correctly inspected good part: When we have k ok-events, the probability that all of them are qualified correctly is $q^k_{ok\to ok}$. However, the actual number may vary statistically. The expectation rate for a correct inspection of all parts is found by summing up the cases for all possible values of k, considering the corresponding probability of occurrence $P_P(k)$:

$$Q_{OK\to OK} = \sum_{k=0}^{\infty} P_P(k) \cdot q^k_{ok\to ok}. \tag{1.2}$$

Such a probability Q_{LM} is denoted *inspection rate* in the following, the indices L an M standing for the inspection outcomes OK and NOK.

False reject: A failure in the inspection of a good part results in false reject. This occurs for any other case than the single case of a correct inspection results, i.e.

$$Q_{OK\to NOK} = 1 - Q_{OK\to OK}. \tag{1.3}$$

Inspecting a bad part

A bad part may carry additional defects (nok-events), together with an arbitrary number k of ok-defects. Normally, e.g. for acceptable production yield, the case of more than one defect per part should be negligible. Even with only one nok-event, there are several cases to be considered.

Correctly inspected bad part: A completely correct inspection result needs the single nok-event and all of the k ok-events to be qualified correctly. Similar to above, the total inspection rate results as

$$Q^{(i)}_{NOK\to NOK} = q_{nok\to nok} \cdot \sum_{k=0}^{\infty} P_P(k) \cdot q^k_{ok\to ok}. \tag{1.4}$$

Seemingly correct inspection of a bad part: Unfortunately, there are two more cases, in which the overall inspection result is correct, while failures happen on the event-level. In the first case, one or more of the ok-events are mis-qualified to be of nok-type. Since the real defect is also correctly qualified, this does not alter the overall outcome. This case happens with a rate of

$$Q^{(ii)}_{NOK\to NOK} = q_{nok\to nok} \cdot \left[1 - \sum_{k=0}^{\infty} P_P(k) \cdot q^k_{ok\to ok}\right]. \tag{1.5}$$

Things may even be worse, when one or more ok-events are falsely qualified to be of nok-type and the one real defect is simultaneously mis-qualified to be ok. Such a double-failure happens with a rate of

$$Q^{(iii)}_{NOK\to NOK} = q_{nok\to ok} \cdot \left[1 - \sum_{k=0}^{\infty} P_P(k) \cdot q^k_{ok\to ok} \right] \tag{1.6}$$

Even though the two last cases rarely happen, it is worth to keep an eye on these numbers as they indicate that something is really going wrong during inspection.

False accept: Finally, a bad part will falsely be accepted to be OK, when all the ok-events are correctly qualified as such and, in addition, the one nok-event is mis-qualified to also be of ok-type. This occurs with an inspection rate of

$$Q_{NOK\to OK} = q_{nok\to ok} \cdot \sum_{k=0}^{\infty} P_P(k) \cdot q^k_{(ok\to ok)}. \tag{1.7}$$

Frequency of good and bad parts

So far, we have regarded the inspection of a given good part or a given bad part. The real production, however, contains both. The relative fraction of good and bad parts can be derived from what we already know. The average number of all defects results as sum over all nok-events $h_{nok} = \sum_{l\in nok} h_l$. The probability that a part under test carries m nok-events is again described by a Poisson distribution $P_P(m)$ with the expected value $h = h_{nok}$.

Consequently, good parts occur with a probability of $P_{OK} = P_P(m = 0)$, and bad parts cover all other cases $P_{NOK} = 1 - P_{OK}$. With these occurrence probabilities, we arrive at weighted inspection rates $Q^\star_{LM} = P_L \cdot Q_{LM}$, describing the frequencies at which correct and false inspection happens for the actual production situation.

3 Application

3.1 Scenario modeling

For illustration we take an example, for which the importance of correct inspection is obvious: quality control of pharma packaging like

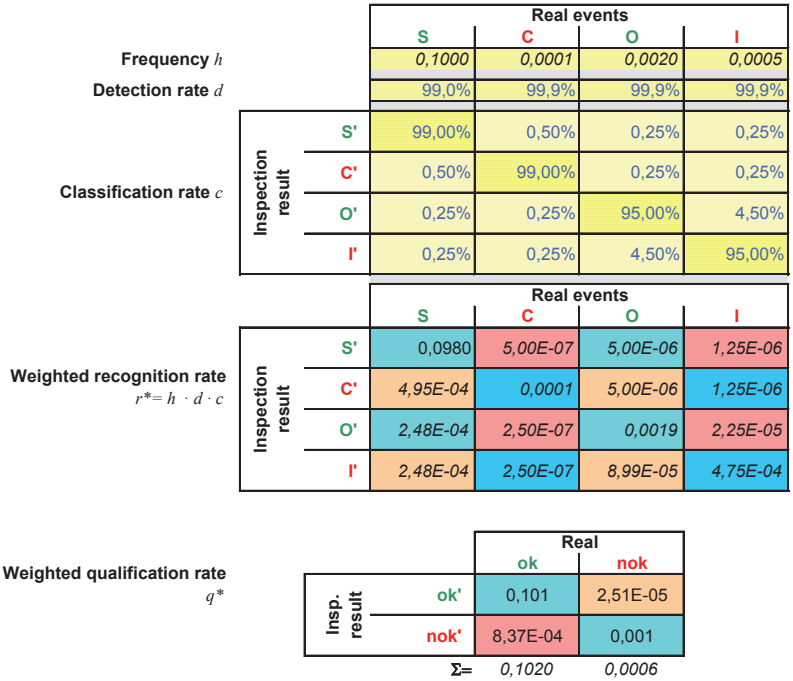


Figure 1.1: Spread sheet calculation of event-based inspection process for the example of a vial inspection. Yellow fields correspond to input parameters.

vials. For simplicity, we consider only four types of events: cracks (C), scratches (S), contamination on the outer (O), and on the inner surface (I) of the vial. The scheme developed in the preceding section can completely be implemented in a spread sheet. The stage related to individual events is shown in figure 1.1. The top row represents the events, their color stands for the quality class. For the present customer, scratches and contaminations on the outside are still acceptable (ok-events, in green), while a contamination inside the vial and cracks are considered to be defects (nok, in red).

The next row contains the average number h_i of events per vial: in average every tenth has a scratch, one out of 10000 is cracked, one out

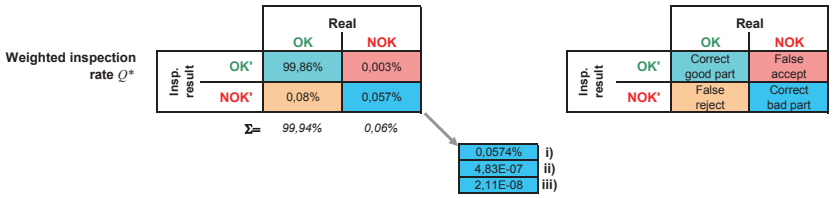


Figure 1.2: Weighted inspection rates $Q^{\star}_{LM}$ corresponding to the data in fig. 1.1. The meaning of each respective field is given in the r.h.s. table. Also, the cases i)...iii) of (seemingly) correct NOK-inspection are indicated.

of 500 has a contamination on the outside, and a contamination inside occurs every 2000 vials. In the row below, the detection rates d_i are given for each event type. In this example we have assumed all detection rates to be 99.9%, except for the hardly visible scratches with $d_S = 90\%$.

The detected events are then classified, and the classification rate is cross-tabulated in a 4×4 table, comparing real defect type (columns) with the inspection outcome (rows, letters with apostrophe). Such tables are known in similar fields of statistics as confusion matrices or contingency tables [6, 7, 11]. Here, the values are normalized such that each column sums up to 100%. This corresponds to the viewpoint that each detected event will be classified into one of the given classes. Ideally, each field of the diagonal d_{ii} should be close to 100%. Here we have assumed lower values.

The confusion table below shows the weighted recognition rates $r^{\star}_{ij} = h_i \cdot d_i \cdot c_{ij}$. The colors chosen for this table indicate to which kind of (mis-)qualification the field belongs. Accordingly, corresponding field entries are summed up to give the 2×2 qualification table, representing the qualification rate $q^{\star}_{lm}$.

The inspection process chain is continued in figure 1.2 with the step to the quality of the whole parts under test. The table shows the weighted inspection rates $Q^{\star}_{LM}$. All fields sum up to 100%, e.g. to the whole production.

The tables in figures 1.1 and 1.2 could for instance be the result of searching the inspection performance (d_i, c_{ij}) necessary to achieve acceptable false-accept and false-reject rates for a given production status

(h_i). In other words, the scheme provides the opportunity to derive quantitative specifications for an inspection system, required to be capable for a given application. Note that the choice of d_i and c_{ij} is not unique, bad detection or classification of one type of defect may be balanced with better performance on other defects.

The scheme is also helpful in calculating scenarios, when conditions change. Let us assume that a modification of a picker arm causes an increase in occurrence of visual contaminations on the vials outside, here from $h_o = 0.002$ to 0.2. As consequence, the false-reject rate increases to an unacceptable value of 1.14% of the whole production. The reason is that contaminations on the inside and on the outside of the vial look similar. This is reflected in relatively large cross-classification rates - which before had not been identified to be of potential harm.

As first consequence, the inspection system manufacturer is forced to increase the classifier performance. This will have natural limits as long as the image features of both events are similar. Only if the cross-classification rates between inner and outer contaminations could be reduced to 0.5%, the false reject rate would go back to around 0.2%. In other words, this problem could require a different illumination and imaging-hardware, in order to achieve sufficient discrimination. If the corresponding effort is too high, the only solution is to improve the production process, i.e. the handling mechanism in order to reduce the contaminations.

Another scenario is the change of the quality rules. Assume, we have a new customer in Far East, who has a strict quality management. He does not accept visual scratches, which in his eyes may indicate hidden damages of the vials. Applying the new quality criteria by simply setting scratches to be of nok-type, causes the false-reject rate to rise to 9.57%. If the new customer is paying well, this might be bearable. However, simultaneously the false-accept rate increases to an unacceptable value of 0.12%. The main reason for the high false-accept rate lies in the relatively low detection rate $d_S = 90\%$ for the hardly visual scratches. Under the new criteria, the cross-classification between scratches and cracks is of no importance, since both are considered to be nok.

False-accept rate and false-reject rate have been taken as independent parameters up to now. It is possible to further reduce complexity by introducing a single risk- or cost measure. The cost of false-reject can e.g. be estimated by lost production cost and the cost of false-accept by

cost of potential customer complaints. Multiplying the corresponding cost per vial with the false-accept and false-reject rates, one arrives at a single cost value as criterion for the system optimization.

3.2 System validation

Confidence

Up to now we have assumed to know the detection- and classification rates of the inspection system. For an existing system to be validated, in particular in the situation of acceptance, these numbers need to be determined from reality. The only practical way to do so is empirically by using samples and counting the detected and classified events.

Estimating a rate $p = k/N$ from the number of "hits" k in a sample of size N is connected with some uncertainty Δp. Since in our case, the relevant rates are either close to 1 (detection rate, classification rate c_{ii}) or close to 0 (mis-detection rate $1 - d_i$, cross-detection rate c_{ij} for $i \neq j$), these uncertainties need to be very small.

The statistical uncertainty is typically described by the confidence interval, in which the real rate will lie (with a certain confidence level $1 - \alpha$) [6]. In our case, the underlying statistics is the Binomial distribution. In literature, a variety of (approximations for the) confidence intervals for this distribution can be found [12,13]. A good compromise between simplicity and accuracy is the following approximation:

$$\Delta p \simeq z_{1-\alpha/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{N}}}, \tag{1.8}$$

with $\tilde{p} = \tilde{k}/\tilde{N}$, $\tilde{k} = k + 2$, and $\tilde{N} = N + 4$, and $z_{1-\alpha/2}$ being the corresponding quantile of the normal distribution.

When we decide for a fixed value of Δp, the necessary sample size can under certain conditions be approximated to be

$$N \simeq z^2_{1-\alpha/2} \cdot \frac{p(1-p)}{\Delta p} \simeq z^2_{1-\alpha/2} \cdot \frac{p}{\Delta p} \tag{1.9}$$

The approximation is valid for $N \gg 1$ and small rates $p \ll 1$; a similar approximation holds for large rates $q = 1 - p \simeq 1$.

If we pragmatically define a relative uncertainty $\Delta p/p$ and require this fraction not to exceed a certain limit $c \geq \Delta p/p$, we are led to an estimation of the minimum number of necessary hits of $k \geq z^2_{1-\alpha/2}/c^2$. For a typically confidence interval of 95% and a choice of $c = 1/10$, we find that $k \geq 400$ hits are necessary to determine the searched rate with acceptable reliability. This, however, is just a rough approximation to be verified for the particular case.

For instance, if we were interested in the detection rate d_i, it makes sense to look at the undetected events (="hits") and calculate the detection rate $d = 1 - p = 1 - k/N$. To achieve the estimated reliability, we need to run the inspection system and check its results, until $k = 400$ undetected events are found. It may be advisable to do this with samples that have been collected before and not to wait in running production for enough (undetected) events. Also, one can use parts, carrying many events. However, it is important that these samples reflects the actual production and the features of the events are similar.

Equation (1.8) describes a double-sided confidence interval. This choice may lead to disputes between system supplier and user in the acceptance, with different viewpoint on when a specified limit value is achieved. Alternatively, one can use one-sided intervals by using the quantile $z_{1-\alpha}$ [13].

Reference validity

We have implicitly assumed to know the real properties of any event, when calculating the detection- and classification rates. In general, there is some kind of reference to which the inspection system is compared, for example a human inspector or a precision instrument. Obviously, in this comparison the automatic system can never be better than the reference, for which we have to consider that it might not always be perfect. From there it is vital to use a reliable reference (e.g. use microscopes instead of the bare eye) and check the reliability. It might be worth to carefully re-check cases of disagreement between automatic system and reference judgment [11]. Please note that the application of the scheme presented here is not limited to automatic systems.

The assumption of a reliable reference statement, that is underlying the present approach, is in so far optimistic. There are other approaches [4, 8], which do in fact compare uncertain classification decisions with

another. Such an approach may open room for further improvements of the present scheme.

In using a reference system, one should also consider that there are events that are correctly recognized by the system, but may have disappeared when the sample is presented to the reference inspection afterwards. Typical examples are dust particles or liquid droplets. A different case is, when the system triggers, while there is no corresponding event in reality (*pseudo-defects*). The cause may lie in camera noise or in cosmic radiation. Such pseudo-defects can be included in the scheme presented here - with the only peculiarity that their frequency of occurrence does not depend on the production, but on the inspection system.

Golden Samples and Limit Samples

Classification in the sense of pattern recognition relies on selected image features. On real-world products, these features can easily change, e.g. due to alterations of the production process, like wearing of tools. Hence, the classification performance is very sensitive to such variations and it is therefore advisable to perform all system validations with samples from the actual production. Another reason is the possible occurrence of new events, which have not been considered during design or teaching of the inspection process.

This makes already clear, where the drawbacks are in using Golden Samples for validation. Their only advantage lies in simplicity and reduced effort. Still, this may in practice be valid reasons to use Golden Samples, however, one should be aware of the fact that using them will never yield the performance of the inspection process for the actual production state. Golden Samples may still be valuable for basic functionality tests and repeatability tests.

Limit Samples appear to be better suited for the determination of detection and classification performance . These would be samples, which are close to the detection limit (e.g. showing low image contrast) or close to classification limits (having image features which are close to the classifiers class borders). But detection-, and in particular classification limits may change with every teaching of the inspection system. Moreover, determining the system behavior with Limit Samples does not indicate, which effect the actual system setting has on the running production - as long as it is not clear how close or how far the events occurring in real production are from the detection and classification limits.

Continuous measures

The presented scheme can only be applied to classifying systems. However, in reality continuous measures are often also used as quality criteria. A typical example is the geometrical extension of a defect. There is, however, a way to include continuous measures at least approximately by using bins. In the case of the defect size one can for instance define several size classes, which could conform to the quality criteria. A binning in size classes also helps to reduce another disadvantage of the scheme, because the size-dependencies of the detection- and classification rates have been ignored completely to keep the scheme simple. A size dependency of the detection rate may be analyzed separately by determining the Probability of Detection - POD [14].

4 Conclusion

The scheme presented here provides a quantitative model for classifying inspection systems, as used in quality control. The analysis makes clear that the performance of the complete inspection process is determined not only by the performance of the inspection system itself. The production status, i.e. the frequency of the defects, and the quality criteria applied are also significant. Consequently, care has to be taken that all parameters are taken into consideration, when such a system is designed or specified.

The scheme can be realized in a usual spread sheet, allowing for easy use, e.g. in order to compute scenarios and to derive quantitative specifications for the inspection system. It has also been discussed how the corresponding figures can be verified empirically and which complications might arise.

The work presented is intended to provide common background for help both suppliers and users, helping to agree on validation procedures and acceptance criteria in an early project stage.

References

1. *DIN 1319: Fundamentals of metrology*, 1995.
2. *DIN V ENV 13005: GUM (Guide to the Expression of Uncertainty in Measurements*, 1999.
3. E. Dietrich, A. Schulze, S. Conrad, *Eignungsnachweis von Messsystemen*. München: Hanser Verlag, 2008.
4. *AIAG: Measurement System Analysis*, Automotive Industry Action Group, Southfield, USA, 2002.
5. *ISO 22514-7: Statistical methods in process management - Capability and performance - Part 7, Capability of measurement processes*, 2013.
6. J. Hartung, *Statistik*. München: Oldenbourg Verlag, 2009.
7. A. Agresti, *An Introduction to Categorical Data Analysis*. Hoboken: JohnWiley and Sons, 2007.
8. W. N. van Wieringen, E. R. van den Heuvel, "A comparison of methods for the evaluation of binary measurement systems," *Quality Engineering*, vol. 17, pp. 495–507, 2005.
9. R. Neubecker, "Fähigkeitsbewertung klassifizierender Bildverarbeitungssysteme für Prüfaufgaben - Teil 1," *Technisches Messen*, vol. 81, no. 9, pp. 422–430, 2014.
10. ——, "Fähigkeitsbewertung klassifizierender Bildverarbeitungssysteme für Prüfaufgaben - Teil 2," *Technisches Messen*, vol. 81, no. 10, pp. 499–510, 2014.
11. G. M. Foody, "Status of land cover classification accuracy assessment," *Remote Sens. Environ.*, vol. 80, pp. 185–201, 2002.
12. L. D. Brown, T. T. Cal, A. DasGupta, "Interval estimation for a binomial proportion," *Statistical Science*, vol. 16, pp. 101–133, 2001.
13. "Nist engineering statistics handbook: Confidence intervals," URL: http://www.itl.nist.gov/div898/handbook/prc/section2/prc241.htm [access 31.1.2014], 2013.
14. "MIL-HDBK-1823A: Department of Defense Handbook: Nondestructive Evaluation (NDE) System, Reliability Assessment," USA, 2009.

Struktur zur Auswahl und Implementierung von bildverarbeitenden Systemen in der Fertigungsautomation

Alexander Grote und Erwin Schwab

Fachhochschule Südwestfalen, Fachbereich Maschinenbau,
Frauenstuhlweg 31, 58644 Iserlohn

Zusammenfassung Eine Problematik bei der Integration von Bildverarbeitungssystemen in der Produktionstechnik ergibt sich aus der Vielzahl möglicher Ansätze, Aufgabenstellungen der Produktionsautomation mittels Bildverarbeitung zu lösen. Die Auswahl des bildgebenden Systems als Datenquelle für das Bildverarbeitungssystem stellt dabei die Weichen für die weitere Projektierung und Gestaltung des Bildverarbeitungssystems und dessen Integration in die Produktionsanlage. Dieser Beitrag soll eine kurze Übersicht zu bereits bestehenden, erfolgreichen Implementierungen von 3D-Bildverarbeitungssystemen in der Produktionstechnik geben und dabei genauer das methodische Vorgehen bei der Projektierung dieser Systeme beschreiben. Es wird ein methodisches Vorgehen definiert, das es einem Entwickler von Produktionssystemen ermöglicht, die notwendigen Maßnahmen zur Realisierung eines effizienten und effektiven 3D-Bildverarbeitungssystems im Verbund mit dem Produktionssystem zu definieren.

1 Einleitung

Die stetig fortschreitende Entwicklung neuer Technologien verkürzt die Innovationszyklen immer mehr. Der Wandel vom Verkäufermarkt hin zum Käufermarkt erfordert immer mehr die Entwicklung von kundenindividueller Lösungen. Eine Folge sind kürzere Marktzyklen, daraus resultiert der Bedarf an häufigen Neuplanungen von Fertigungsprozessen oder deren Rekonfigurationen (vgl. [1]).

Dabei wird der Fertigungsprozess durch eine ständig fortschreitende Automatisierung und immer höhere Anforderungen an die Fertigungsqualität charakterisiert. Dieser Fortschritt wird teilweise erst durch den Einsatz der Querschnittstechnologie Bildverarbeitung (BV) ermöglicht. Dieser Beitrag und die beschriebene Struktur befassen sich primär mit der industriellen Bildverarbeitung, sowie der 3D-Bildgebung und deren Anwendung in der Produktionstechnik. Es wird die Entwicklung geeigneter Produktionsanlagen mit 3D-Bildverarbeitungsystemen (BVS) oder die nachträgliche Implementierung eines solchen Systems in bestehende Produktionsanlagen betrachtet.

Anhand von Beispielapplikationen wird die Struktur, die zur Auswahl der jeweiligen bildgebenden Systeme geführt hat, beschrieben. Dabei wird die Vielfalt der 3D-Bildverarbeitungsmethoden als Problemgröße bei der Entwicklung entsprechender Systeme erläutert. Es wird auch das Spektrum der Formen und Größen und die Losgröße der Produkte, die in Anlagen mittels 3D-Bildverarbeitung erfasst werden, betrachtet. Die Ausdrucksweise dieses Beitrags richtet sich nach den in [2] definierten Begriffen, um allgemein verständlich zu sein.

2 Beispiel implementierter Systeme

Im Folgenden werden drei Beispiele für bereits bestehende 3D-BVS gegeben, anhand derer die Merkmale der entwickelten Struktur beschrieben werden kann. Die beschriebenen Systeme sind in Ihren Komponenten teilweise recht ähnlich, in der Funktionalität jedoch sehr unterschiedlich. Zum einen sind es Systeme, die im Nachhinein in den Fertigungsprozess eingebracht wurden, zum anderen BVS, die mit dem Fertigungsprozess zusammen entstanden sind.

Das erste System erfasst mit Hilfe eines Lichtschnittsensors, die zu untersuchenden Merkmale. Mit dem System werden die relevanten Merkmale von Spannbetonbauteilen in der Produktion auf ihre Maßhaltigkeit und ihren Bezug zueinander geprüft und elektronisch aufgezeichnet.

Die Objekteigenschaften und die Art der Merkmale verlangen ein System, das auf der einen Seite eine möglichst gute Auflösung des einzelnen Merkmals gewährleistet, aber auch die Prüfung der Maßhaltigkeit der Bauteilgeometrie über die gesamte Länge des Objektes zulässt.

Eine Digitalisierung des gesamten Objektes, mittels eines Sensors ist mit der geforderten Taktzeit des Produktionsprozesses nicht vereinbar. So wurde bei den bildgebenden Systemen darauf geachtet, dass die für die Vermessung nötigen Merkmale innerhalb der zeitlichen Vorgaben möglichst genau erfasst werden.

Bei dem zu vermessenden Objekten, in Abbildung 2.1 dargestellt, handelt es sich um Bahnschwellen aus Spannbeton. Mit dem BVS werden die relevanten Maße der Schienenauflager erfasst und dem Leitprozess übergeben. Die Merkmale, die für die Vermessung notwendig

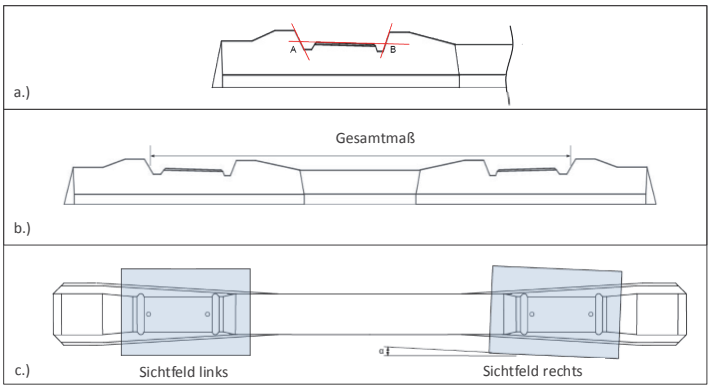


Abbildung 2.1: Objekt und zu ermittelnde Maße, a.) Auflager mit Kennzeichnung des relevanten Maßes, b.) Kennzeichnung Gesamtmaß, c.) schematische Darstellung der Szene mit Sichtfeld.

sind, liegen innerhalb des sogenannten Auflagers. Dies führte zu der BV-Strategie, dass mit je einem Lichtschnittsensor ein Auflager digitalisiert wird. Die Vermessung erfolgt synchron, indem die Lichtschnittsensoren synchron über die Auflager geführt werden, wobei die beiden Sensoren zueinander kalibriert sind, um das Gesamtmaß (siehe 2.1 b.)) zu erfassen.

Diese Anordnung der Sensoren ermöglicht die zeitunkritische Erfassung des Objektes, erschwert jedoch die Erfassung des Gesamtmaßes, da die beiden Sensoren keinen gemeinsamen Bildbereich erzeugen (vgl. 2.1 c.)). Eine detaillierte Beschreibung des Aufbaus ist [3] zu entnehmen.

Das zweite System, ist eine Robot-Vision Anwendung, bei der zylindrisch geformte Bauteile aus Ladungsträgern mittels 6-Achsen Industrieroboter gegriffen werden und so dem Fertigungsprozess bereitgestellt werden. Das zugehörige BVS ermittelt die Greifpunkte, indem mit dem Lichtschnittverfahren die Ladungsträger digitalisiert werden und mittels einer adaptiven Software der Greifpunkt auf dem jeweiligen Objekt errechnet wird.

Der Prozess, in den das BVS integriert wurde, erfordert dass zylinderförmige Bauteile unterschiedlicher Längen von 300-1800 mm und unterschiedliche Durchmesser von 40-300mm handhabbar seien müssen. Die nachfolgende Abbildung zeigt eine schematische Darstellung der Szene und des anlagentechnischen Bereichs zur Digitalisierung dieser. Der zu ermittelnde Greifpunkt liegt auf dem Scheitel des

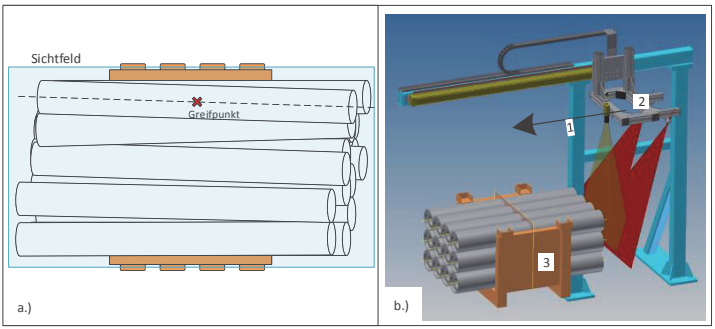


Abbildung 2.2: a.) schematische Darstellung der Szene mit Sichtfeld und zu ermittelndem Greifpunkt b.) Anlagentechnischer Teil des BVS bestehend aus: dem Schlitten mit Sensorik (2) zur Digitalisierung der Szene (3), Bahn des Lichtschnittsensors entlang (1).

zylindrischen Körpers und dem Mittelpunkt definiert durch die Länge des Körpers (vgl. 2.2 a.)). Diese Art der Digitalisierung bietet sich für die konvexe Form der Objekte als ideal an, da die quasi senkrecht zur Achse stehende Laserlinie eine optimale Erfassung der Oberfläche gewährleistet.

Das dritte System ist ebenfalls eine Robot-Vision Anwendung, hier muss die Einbringposition eines Bauteils in ein Galvanikgestell indivi-

duell für jedes Bauteil mittels BV ermittelt werden. Der Sensor muss das Gestell und die Einbringposition erfassen und die Koordinaten für das Einbringen des Bauteils in das Gestell ermitteln.

Die Lokalisierung der Einbringposition erfolgt mittels Stereokamerasystem, das an der sechsten Achse eines 6-Achsen Industrieroboters, der auch das Handling der Bauteile übernimmt, angebracht ist. Zur Digitalisierung erfolgt eine Positionierung des Sensors in eine Beobachtungsposition, so dass der Teilbereich des Gestells, in dem die nächsten Bauteile appliziert werden sollen, erfasst werden kann. Nach der Bildaufnahme werden die Positionen der Haken ermittelt und die Position des optimal zu bestückende Hakens wird an den Roboter übergeben. Der Anwendungsbereich für die Bestückung von Galvanikgestellen ist sehr weit, da die Form der Bauteile so vielseitig wie das Material ist, da heutzutage alle gängigen Grundwerkstoffe aus Metall sowie die meisten bekannten Kunststoffe beschichtet werden können.

Das hier beschriebene System zielt auf galvanische Prozesse, die sehr variantenreich sind und auch durch einen hohen mechanischen Aufwand nicht automatisiert werden können. Dabei muss primär darauf geachtet werden, das das BVS inklusive Sensor mit einer hohen Varianz an Objekten, hier unterschiedliche Gestellformen, umgehen kann. Außerdem ist in dem Prozess vorgesehen, dass neue Varianten auch ohne Expertenwissen in die Produktion mit aufgenommen werden können. Die nachfolgende Abbildung zeigt einen typischen Aufbau, bestehend aus Roboter, Sensor, Gestell und Abholposition.

Ausgehend von den zuvor beschriebenen BVS werden im Folgenden die für die Entwicklung und Entscheidung wichtigen Eigenschaften der BVS zusammengefasst und hinsichtlich ihrer Übertragbarkeit auf die Anwendung in der allgemeinen Projektierungsphase von BVS geprüft.

Die ersten beiden Systeme unterscheiden sich deutlich in der Aufgabe, die sie im Gesamtprozess erfüllen, arbeiten jedoch mit dem gleichen bildgebenden Verfahren. Dementgegen steht der gleiche Aufgabenbereich des dritten Systems mit dem des zweiten Systems, hier ist jedoch ein erheblicher Unterschied in der Bildgebung festzustellen, die Differenz der Systeme setzt sich in den angewandten Methoden der BV fort. Betrachtet man die Merkmale der Systeme getrennt voneinander, kann man deren wesentlichen Merkmale nach den Objekteigenschaften, Aufgaben der Bildverarbeitung und Kennzeichen des Fertigungsprozesses gliedern. Im ersten System werden schwach texturierte Bau-

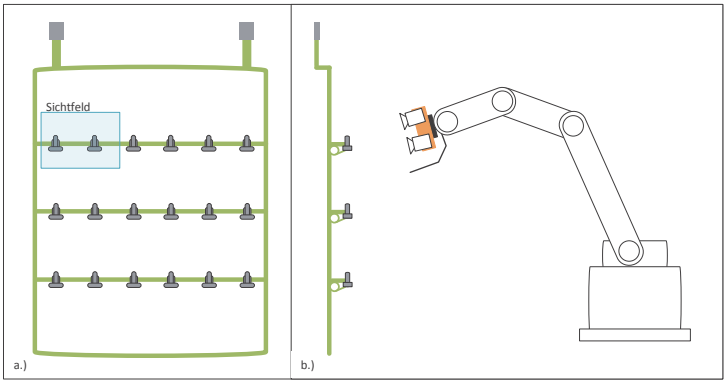


Abbildung 2.3: a.) schematische Darstellung der Szene mit Sichtfeld b.) Roboter in Beobachtungsposition mit Stereokamera zur Digitalisierung der Szene.

teile aus Beton vermessen. Die Objektgeometrie und die Führung des Sensors ermöglichen eine Erfassung des Objektes ohne Abschattungen. Die Aufgabe der BV besteht in der Extraktion und Berechnung der beschriebenen Maße, also einer geometrischen Betrachtung des Objektes mit einer hohen Genauigkeit, wobei auf einfache Methoden der Bildverarbeitung zurückgegriffen werden muss, um diese über den Schnitt von unterschiedlichen Ebenen zu bestimmen. Die mechanischen Gegebenheiten beschränken die Führung des Sensors nicht. Die Taktzeit bestimmt die Digitalisierung mit zwei Sensoren. Die Sensoren und die Digitalisierung sowie die geringe Varianz der zu vermessenden Typen ermöglichen eine relativ einfache Strategie bei dem Vorgehen in der BV. Der Sensor und die Digitalisierung bieten den Vorteil, dass die Szene im Wesentlichen nur aus den zu extrahierenden Merkmalen und kaum Hintergrund besteht. Der Anteil der Bildvorverarbeitung und der Objektsuche fallen sehr gering aus.

Im zweiten System muss der Greifpunkt von schwach texturierten, teilweise reflektierenden, Bauteilen in einem Ladungsträger bestimmt werden. Bei der gewählten Anordnung der Sensoren und ihres Verfahrweges ergeben sich nur wenige Abschattungen. Die Varianz der Objekte, die so gehandhabt werden sollen, ist sehr groß. Die Aufgabe der

Bildverarbeitung besteht in der Extrapolation der Greifpunkte wie oben beschrieben. Dies erfolgt im Wesentlichen über Lage und Geometrie des Objektes. Das BVS muss in der Taktzeit des Fertigungsprozesses arbeiten. Aus den beschriebenen Merkmale und der Varianz an zu erfassenden Objekte ergab sich die BV Strategie:

- Analyse der Szene und Segmentierung (Trennung Objekt Hintergrund)
- Merkmalsextraktion (Bestimmung des Greifpunktes und der Greifbarkeit)

Die Analyse und Extraktion der geforderten Merkmale sind deutlich aufwendiger, als im ersten System. Hinzu kommt, dass eine hohe Varianz an Bauteilen mit diesem BVS behandelt werden muss. Dies führte zusätzlich zur Erhöhung des Aufwandes bei der Softwareentwicklung bei dem BVS, da bei der gesamten BV-Kette [4] und den angewandten Methoden stets auf die Adaption geachtet werden musste. Aufgrund der möglichen Vielfalt an Objekten und Gestellformen beim dritten System, kann zunächst keine eindeutige Aussage zu der Form oder den optischen Eigenschaften der Szene bzw. des Objektes getroffen werden. Einzige Einschränkung ist die Größe des Objekts. Diese muss in das Sichtfeld des Sensors passen. Aufgrund der Formvielfalt, der räumlichen Bedingungen, des umgebenden Prozesses sowie der Taktzeitabhängigkeit konnte von vornherein das Lichtschnittverfahren ausgeschlossen werden. Die sich aus dem Fertigungsprozess ergebenden Eckpunkte wie Objektvarianz und Taktzeit, sowie das Stereokamerasystem definieren die Rahmenbedingungen des BVS. Das Stereokamerasystem wurde aufgrund der Vielseitigkeit der anwendbaren Methoden, die in [5] eingehend erklärt werden, ausgewählt. Die geforderte einfache Erweiterbarkeit setzt die Integration einer umfangreichen Mensch - Maschine - Schnittstelle und entsprechend adaptive Methoden der Bildverarbeitung voraus. So wurde, das Prinzip des geometrischen Matching, vgl. hierzu [6], das vielfach in 2D-Bildverarbeitungssystemen integriert und etabliert ist auf das Stereokamerasystem übertragen. Der Ablauf der Bildverarbeitung verdeutlicht die Vorteile:

1. Matching-Durchlauf linkes und rechtes Kamerabild
2. Bestimmung der korrespondierenden Punkte beider Matchingergebnisse

3. Berechnung der Koordinaten des Korrespondenzpaares

Diese Koordinaten werden als Einbringposition verifiziert und dem Roboter übergeben. Dieses Vorgehen bietet den Vorteil, dass keine 3D Rekonstruktion der Szene erfolgen muss und nur die nötigen Korrespondenzpunkte des Matchingergebnisses trianguliert werden. Durch diese Methodik wird eine minimale Zykluszeit des BVS gewährleistet und es ist gleichzeitig möglich, die einfache Erweiterbarkeit des BVS mit Methoden der Mustererkennung zu gewährleisten.

Anhand des beschriebenen Vorgehens wird deutlich, dass die Auswahl des Sensors das gesamte Automatisierungsprojekt beeinflusst. So kann die Sensorik die Softwareflexibilität einschränken oder den Aufwand bei der Entwicklung der entsprechenden BV-Software erhöhen. Das nötige Wissen zur Entwicklung und Planung des Gesamtsystems ist einmal abhängig von dem Sensor, aber auch von den geforderten Funktionen der BV (Erweiterbarkeit, etc.). Die beschriebenen Applikationen zeigen, dass die Entwicklung von 3D-BVS durch die Merkmale wie in Abbildung 2.4 definiert ist. Anhand dieser Kriterien bietet es sich

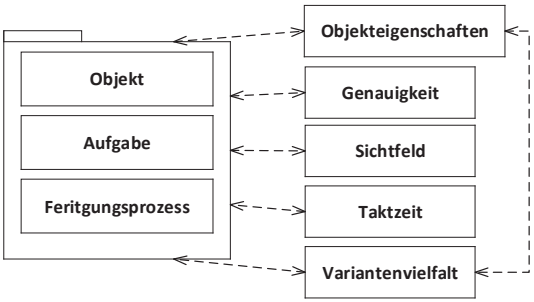


Abbildung 2.4: Einfluss der Merkmale auf die Kriterien des BVS.

an, das bildgebende System auszuwählen. Der ausgewählte Sensor und die Kriterien beschreiben den zu erwartenden Aufwand für die Softwareentwicklung sowie den Gesamtaufwand für die Umsetzung und Implementierung des BVS. Bei dem Ablauf der Planung wurde das Schema aus [7] für die Planung und Durchführung von 3D-Messaufgaben an die Anforderungen der Fertigungstechnik und industriellen BV an-

gepasst. Daraus ergibt sich für die Projektierung als Definition des BVS das Paket wie in Abbildung 2.5 dargestellt. Der Sensors sollte so ausgewählt werden, dass der Aufwand zur Extraktion der gewünschten Merkmale über die Varianz der Objekte minimal ist. Der hier vereinfachte Projektierungsansatz wird in eine Struktur umgesetzt, die die Entwicklung von BVS durch Aufwandsabschätzung und Wissensmanagement unterstützt.

Der wesentliche Unterschied bei der Projektierung des BVS im Gegensatz zur reinen Betrachtung der technischen Machbarkeit ist, den Faktor der Wirtschaftlichkeit mit einzubeziehen. So ist die beste technische Lösung nicht immer die Lösung die im Fertigungsprozess Anwendung findet, da der Fertigungsprozess den Regeln einer wirtschaftlichen Produktion unterliegt. Die Produktion also auch die Produktionsmittel müssen wirtschaftlich bleiben. Außerdem spielt die Zeit, z.B. Time to market eine Rolle, das heißt, es muss auch eine zügige Entwicklung und Inbetriebnahme der Produktionsmittel stattfinden. Aufgrund

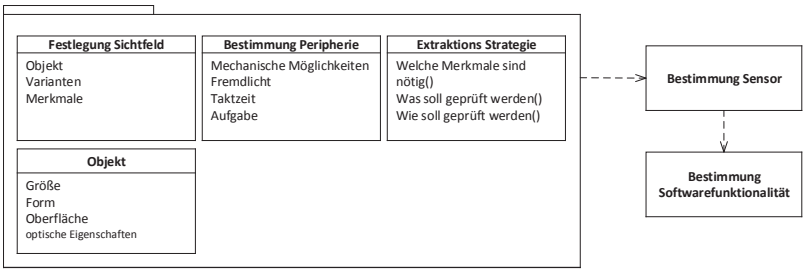


Abbildung 2.5: Strukturdiagramm mit Sicht auf die modellierten Klassen zur Definition eins BVS in der Projektierung, Zielsetzung der zu entwickelnden Softwarestruktur.

der Vielfältigkeit der Methoden zur Lösung eines Problems mittels BV sollte stets nur das Nötigste betrachtet werden und nicht das technisch Mögliche. So hätte beispielsweise bei dem ersten System aufgrund der vorliegenden 3D Digitalisierung des Objektes auch ein Abgleich mit den CAD Daten der Objekte stattfinden können, um die geforderten Merkmale zu erfassen, dies hätte jedoch den Softwareaufwand deutlich gesteigert. Hier gilt es wie in [2] Blatt 2 eine gemeinsame Kommunika-

tion zu nutzen und Anforderungen deutlich zu erfassen und im Sinne der Wirtschaftlichkeit abzuwägen.

3 Ausblick

In [8] werden als Anforderungen der Industrie an die Automatisierungstechnik unter anderem die flexible Automation, Standardisierung, automatische Konfiguration und Kostensenkung genannt. Dabei sollen unter anderem Standards geschaffen werden, die es Automatisierungsherstellern ermöglichen, systemkompatible Komponenten zu generieren. Diesen Gedanken gilt es von der allgemeinen Automatisierungstechnik auch auf die Bildverarbeitung zu übertragen.

Doch die 3D-BV ist teilweise noch immer ein Sonderfall der industriellen BV, obwohl auf dem aktuellen Stand der Technik gerade mittels 3D-BVS sehr viele Problemlösungen machbar sind. Eine Standardisierung würde diesem Trend entgegenkommen, da so mehr Übersicht über die BVS generiert würde. Der Bereich der 3D-BV unterliegt einem ständigen Wandel, weshalb es sich nicht unbedingt anbietet, sich auf die Standardisierung der Komponenten festzulegen, sondern eher das strategische Vorgehen bei der Entwicklung eines solchen Systems zu definieren. So kann erreicht werden, dass schon während der Projektierungsphase eine genaue Betrachtung der Bildverarbeitung stattfindet, um Probleme zu umgehen, die bei einer nicht sachgerechten Auswahl entstehen.

Das vorgeschlagene Vorgehen bei der Entwicklung eines BVS wird momentan in eine Struktur umgesetzt, die es ermöglicht, auf Erfahrungen aus vorangegangen BVS zurückzugreifen und diese mit in die Projektierung zu übernehmen. Es soll eine objektorientiertes Vorgehen geschaffen werden, das es ermöglicht anhand von Blöcken ein System zu definieren und anhand dessen die Struktur der nötigen BV Softwarekomponenten abzuleiten.

Es wurden zunächst die Komponenten 2.6 für die Produktions- und Fertigungstechnik abgeleitet. Dabei wird in diesem Beitrag zunächst nur auf das bildgebende System und die Bildverarbeitungssoftware eingegangen. Es gilt weiterhin die Kommunikation der Komponenten untereinander und mit dem Gesamtprozess zu untersuchen. Die einzelnen Komponenten für eine BVS müssen in nachfolgenden Projekten auf ih-

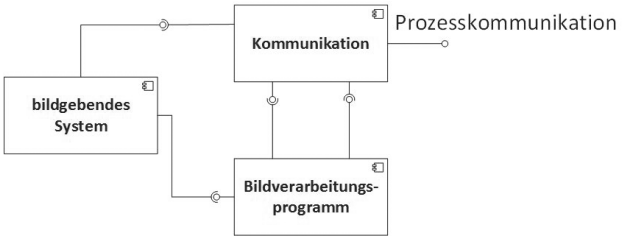


Abbildung 2.6: Komponenten des BVS in UML Notation.

re Merkmale hin geprüft werden und in die Struktur mit eingepflegt werden. Daraus lässt sich eine Gewichtung und Bewertung von Kriterien für diese Struktur entwickeln, die so die Projektierung noch besser unterstützen. Der objektorientierte Aufbau der Struktur kann in die Richtung erweitert werden, dass im Anschluss an die Projektierung ein erster Software Entwurf mit allen nötigen Klassen exportiert werden kann.

Diese starke Verzahnung von Projektierung und Implementierung bietet sich gerade für Maschinenbauer und Systemintgratoren, meist sogenannte KMU, an, da sie nur über ein personell begrenztes Expertenwissen verfügen, sowie durch enge Zeitpläne für die Entwicklung und Inbetriebnahme der Anlagen gebunden sind. Viele Probleme müssen und sollen aus wirtschaftlichen Gründen nicht mehr selber gelöst werden. Es soll vielmehr die vorhandene Problemlösung dem Umsetzer des Automatisierungsprojektes zugänglich gemacht werden. Die vorgestellte Struktur soll Methoden dafür aufzeigen.

Literatur

1. H.-J. Bullinger, H. J. Warnecke und E. Westkämper, Hrsg., *Neue Organisationsformen im Unternehmen*, 2. Aufl., Ser. VDI-Buch. Berlin Heidelberg: Springer, 2003.
2. VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik, Hrsg., *Industrielle Bildverarbeitung*, April 2010, Vol. 01.040.37, 35.240.50, Nr. VDI/VDE 2632.
3. A. Grote und E. Schwab, „Berührungslose optische Vermessung von Spannbetonfertigteilen“, *tm – Technisches Messen*, Vol. 80, 2013.

4. J. Beyerer, C. Frese und F. Puente León, *Automatische Sichtprüfung*. Springer Vieweg, 2012.
5. R. Hartley und A. Zisserman, *Multiple view geometry in computer vision*, 2. Aufl. Cambridge University Press, 2003.
6. C. Demant, A. Springhoff und B. Streicher-Abel, *Industrielle Bildverarbeitung*, 3. Aufl. Springer, 2011.
7. N. Bauer, Hrsg., *Handbuch zur industriellen Bildverarbeitung*, 1. Aufl., Ser. Vision. Fraunhofer-IRB-Verl., 2007.
8. B. Favre-Bulle, *Automatisierung komplexer Industrieprozesse*, 1. Aufl. Springer, 2004.

Self-optimized adaptive algorithm solutions for vision systems

Timur Rashba and Sergei Richter

Math & Tech Engineering GmbH
Robert-Bosch-Str 6/1, 72654, Neckartenzlingen

Abstract Automated development of imaging software is one of the open fields of research and practical applications in computer vision. We present our current developments of the adaptive algorithm approach, based on usage of image processing libraries, optimization and machine learning methods. The concept of the adaptive algorithm approach, its structure, working processes and possible applications are described.

1 Introduction

Accurate, robust and fast analysis of visual information is a key tool for all fields of research and industrial applications. Most of visual information analysis tasks cannot be solved by a single fixed set of image or data processing algorithms. The complete solution requires sophisticated and intelligent combination of methods and approaches, tuning their parameters and extensive statistical analysis and learning. The development of an entire image processing algorithmic solution is still done in large part by human.

Instead we wish to move the human intervention and interaction with the vision system on the next level [1]. The human would need to define the solution on a meta-language level, to provide initial labels (goals) and then to control and to correct the results, see Figure 3.1. Other part of this work, namely building of algorithm networks and their optimization, would be done by the computer system automatically, using user-defined labels (goals), database of algorithms, optimization and data mining methods.

This is possible, because modern and efficient hardware (high precision and fast imaging cameras, fast multi-core and GPU accelerated

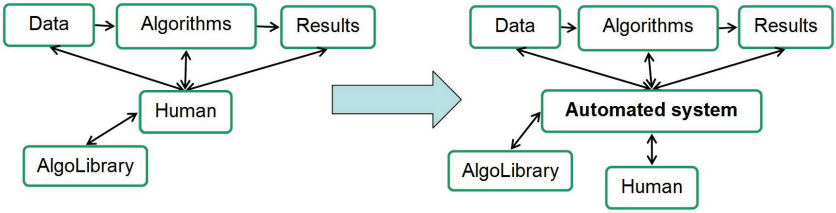


Figure 3.1: Human-vision system interaction: from the direct development to the automated development.

computer systems), vision software (extensive and reach image processing libraries), and mathematical methods (data mining, optimization and machine learning) are well developed and widely available.

In this work we present our current developments of adaptive algorithm approach based on usage of image processing libraries, optimization and machine learning methods. The concept of the adaptive algorithm approach, its structure, working processes and possible applications will be described.

2 State of the art

We will refer shortly to different methods and advances in the field of automated development of the data analysis software, in particular image processing software.

Significant progress in the research of automated software development, in particular, of imaging software, is done during last two decades. These are large variety of evolutionary algorithms [2], in particular, genetic programming [3], sequential parameter optimization [4], machine vision with generic algorithms [5], global optimization algorithms [6], self-configuring applications [7], automatic feature generation [8].

On the other hand generic and widely available automated development of the imaging software is still in the premature stage. Therefore further studies and practical applications of different methods are needed. One of them, adaptive algorithm solution, is presented in this work.

3 Adaptive algorithms

The concept of the adaptive algorithm approach, its structure and working processes will be described in this section.

The adaptive algorithm approach is a method of building algorithm pipeline for a specific image processing task using user-defined data and database of elementary (does not mean simple) algorithms. The role of the image software user or developer is to define the required algorithm pipeline with an abstract meta-language description.

Principal building blocks needed for automated development of image processing software are: knowledge (expert) data, the meta-language description, the goal function, database of building algorithms, the constructor of algorithm pipeline and the optimizer.

- **Knowledge (expert) data (Labels)**
  - Any form of expert knowledge data, e.g. marked regions, human decisions as true/false, ground truth measurement results, labels
  - It is one of the inputs for the goal function, see below.
- **Meta-language description (MetaAlgo)**
  - High level symbolic abstraction of the data processing pipeline,
  - Contains sequential description of the standard image processing steps to be used for a given task. E.g. calibration, finding region of interest, measure; another example: segmentation, feature extraction, classification:

Table 3.1: Sample MetaAlgo description for the surface inspection task.

id	MetaAlgo	InID	InData	OutData
0	Segmentation	Input	Images	Regions
1	Filtering	0	Regions	Regions
2	Features	1	Regions	Features
3	Classification	2	Features	Classes

- **Goal function**
  - A single value which describes the distance (discrepancy) between the algorithm results and the knowledge data,
  - It can be, e.g. accuracy as a statistical measure:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.1}$$

    where T, F, P, N are standard notations for true and false, positives and negatives, respectively.

- **Database of building algorithms (AlgoLibrary)**
  - elementary or complex algorithms from image processing libraries (OpenCV, ITK, VTK, Intel IPP, Halcon),
  - have standard pipeline input-output interfaces,
  - are sorted by type (measuring, segmentation, features, classification, etc), which is related to the meta-language description,
  - stored in AlgoLibrary together with default (and statistically tested) parameters.

- **Constructor of algorithm pipeline (AlgoConstructor)**
  - Meta-language compiler, which builds specific algorithm pipeline, as given in MetaAlgo description, by bringing together corresponding elementary algorithms from AlgoLibrary,
  - Together with algorithms their parameters are also set, either as default values (initially) or as given by the Optimizer.

- **Optimizer**
  - Runs the algorithm pipeline, build by the AlgoConstructor, calculates the Goal value and provides the description for the next algorithm pipeline to be build,
  - Different types of optimization methods can be used, depending on the specific image processing task.

The algorithm pipeline concept is quite similar to the one used in The Insight Toolkit (ITK) [9].

Having building blocks defined above, we have created the software, with the following workflow:

- The solution to a specific visual information analysis problem is formulated using a meta-algorithm language.
- The goal function is defined as a mathematically formulated description of the desired system goal,
- The expert knowledge in the form of user-defined labels and ground truth images is brought to the system.
- The network of adaptive algorithm pipelines for data analysis is self-constructed using building algorithms database.
- The adaptive algorithm network and their parameters are continuously optimized using user-defined labels, new data and algorithm database.
- The obtained results are validated and stored as reference data for verification of the long-term stability of the system.

Schematically the adaptive algorithm workflow is shown in Figure 3.2. There are two principal blocks: supervised and unsupervised optimization blocks. First data enters the Supervised block and form the Dataset. Then they get assigned expert knowledge-labels. Next both data and labels enter the Unsupervised optimization block and get processed by the initial algorithm pipeline. After that the goal value – the distance between obtained results and labels – is measured and sent to the Optimizer for creating next description of the pipeline. The unsupervised optimization continues until it reaches the breaking criteria or it is manually stopped, if the user finds that more input data are needed for further optimization. At the end of the day the optimal algorithm pipeline is created.

Several important issues to be noted and commented. The optimization takes place over different competing algorithm pipelines with different parameters. Therefore we call it adaptive algorithm approach and we see it as a possible way towards automated development of image processing software. It is clear, that the optimization – optimum

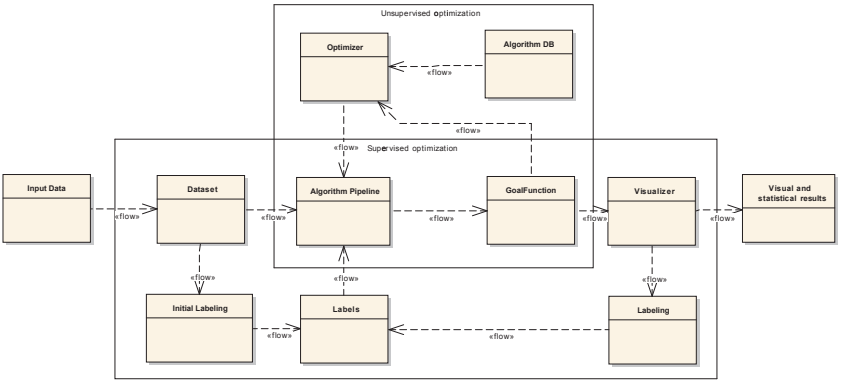


Figure 3.2: Adaptive algorithm approach workflow.

search task – is an ill-posed problem for our case of a multi-parameter and multi-algorithm search spaces. To guarantee the convergence of the optimization search we need the supervised optimization block, which is controlled by the user.

4 Case studies

Typical application and case study of automated development of imaging software is a surface inspection task. Selected images were loaded to the test software. The default algorithm pipeline, described in meta-language (Table 3.1), without classifier, was initially loaded and used for creation of the candidate regions. The regions were labeled by the user. In the first case, the regions with black stripes (top left in Figure 3.3) , in the second case, the regions containing bright spots (bottom left in Figure 3.3) . After that the same algorithm pipeline was loaded with classifier and run into optimization loop. The result of optimization is shown on the right images in Figure 3.3.

5 Summary

The adaptive algorithm approach is described as a way towards automated development of the image processing software. The concept of

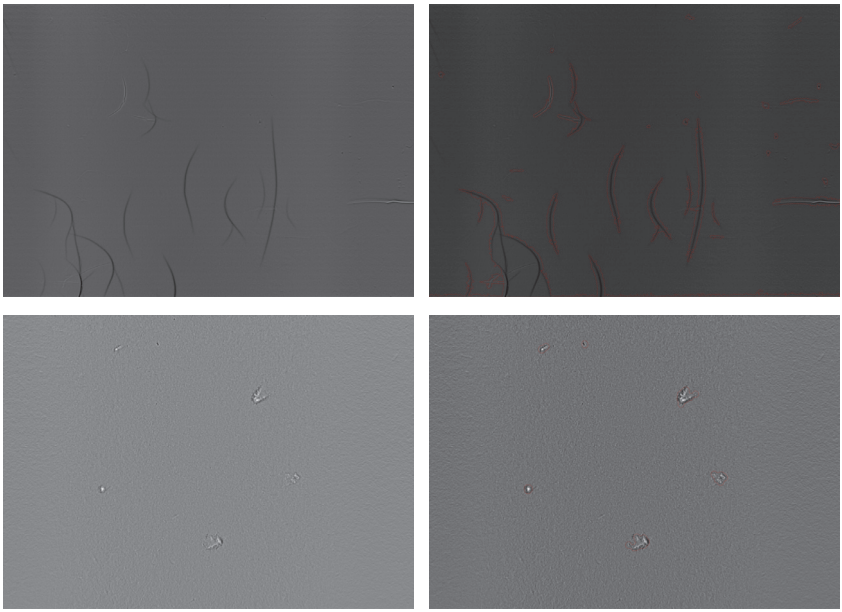


Figure 3.3: Examples of adaptive algorithm results for the plastic (top) and wood (bottom) surface inspection. Left – original image, right – found defects marked as red contours.

the adaptive algorithm approach, its structure and workflow are presented. The important component is a usage of standard image processing libraries (OpenCV, ITK, VTK, Intel IPP, Halcon), which are widely available and can be directly used in automatically generated image processing pipelines. Our further plans include statistical testing of the proposed approach on available labeled image databases.

References

1. S. Richter and T. Rashba, "Automation of the development of imaging software," in *SpectroNet Collaboration Forum 2014*, 2014.

2. T. Bartz-Beielstein, J. Branke, J. Mehnen, and O. Mersmann, "Evolutionary algorithms (pre-peer reviewed version)," this is the pre-peer reviewed ver-

sion of the following article: Bartz-Beielstein, T. and Branke, J. and Mehnen, J. and Mersmann, O.: Evolutionary Algorithms. WIREs Data Mining Knowl Discov 2014, 4:178-195. doi:10.1002/widm.112, 2014.

3. J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA, USA: MIT Press, 1992.
4. T. Bartz-Beielstein, "Research topics in sequential parameter optimization," Presentation — ESF Workshop Rome, 05 2012.
5. U. Köthe, "Reusable software in computer vision," in *Handbook on Computer Vision and Applications*, P. G. B. Jähne, H. Haußecker, Ed. Academic Press, 1999.
6. T. Weise, *Global Optimization Algorithms – Theory and Application*. Germany: it-weise.de (self-published), 2009. [Online]. Available: http://www.it-weise.de/projects/book.pdf
7. M. Hall, Y. Gil, and R. Lucas, "Self-configuring applications for heterogeneous systems: Program composition and optimization using cognitive techniques," in *Proceedings of the IEEE, Special Issue on Cutting-Edge Computing: Using New Commodity Architectures*, vol. 96, no. 5, 2008. [Online]. Available: papers/hall-gil-lucas-ieeeprocs08.pdf
8. W. Konen, "Self-configuration from a Machine-Learning Perspective," *ArXiv e-prints*, May 2011.
9. H. J. Johnson, M. McCormick, L. Ibáñez, and T. I. S. Consortium, *The ITK Software Guide*, 3rd ed., Kitware, Inc., 2013, *In press*.

Realization and evaluation of image processing tasks based on synthetic sensor data: 2 use cases

Stephan Irgenfried, Frank Dittrich and Heinz Wörn

Karlsruhe Institute of Technology KIT,
Institute for Anthropomatics and Robotics IAR
Intelligent Process Control and Robotics Lab IPR
Engler-Bunte-Ring 8, D-76131 Karlsruhe

Abstract In this paper we present two use cases, in which synthetic sensor data is used for selection and training of image segmentation algorithms. The data and the corresponding ground truth information is thereby created using virtual 3D scenes and with the help of computer graphics algorithms. For the realization of both image processing tasks, we used synthetic training data which varies in the creation process, the expressiveness and the scene information type. Based on synthetic and real-world testing data, we show the overall high performance of the approaches, and thereby motivate the applicability of synthetic image data for the engineering of real-world image processing tasks. In addition, this work shows the usage of several different, publically availably tools for creation of synthetic sensor data for real-time and non real-time applicaitons.

1 Introduction

Training and evaluation of image processing algorithms requires a set of images together with the output expected from the system, the ground truth. As stated in [1], training data for the algorithms can be created either by using real sensors or artificially with the help of computer graphics. Using real sensors image data includes natural variations and distortions affecting the image acquisition step, but image content has to be labeled in a manual process, which is time-consuming and subjectively biased. In case using data created from a virtual 3D-scene and sensor simulation, the image content (ground truth) is already known

by design and a large number of images can be created very efficiently. Drawback of this approach is often a lack of realism due to insufficient modeling of objects, surface properties, light sources and sensors.
We focus on the latter and give examples, how to create synthetic data to be used to train and evaluate image processing algorithms. Tools being publically available and having a learning curve managable for computer vision developers which, in most cases, are not computer graphics experts in parallel are used.

2 Related work

This work is mainly influenced by two groups of work done in the area of engineering machine vision systems. The first one goes back to the 1980s and 1990s, where quite a lot of research work was done under the terms C̈AD-based visionänd s̈ensor planning for machine vision.̈ For a good overview on work done until 1995 see [2]. CAD-based vision primarily focused on finding camera positions suitable to solve a given vision task based on geometric considerations and optimizing for feature visibility. By that time, the lack of computing power and availabilty of global illumination rendering algorithms made sensor simulation reaching a high degree of physical realism a vision, which is about to become reality nowadays. The second group of work is selection and benchmarking of computer vision algorithms based on image datasets[1][2], especially synthetically created image data. For design guidelines on how to create such datasets, see [3]. In [4] the pros and cons of using synthetic image data for algorithm benchmarking, especially optical flow algorithms, are discussed.

3 Application

As discussed in [4], synthetic image data can be of use for machine vision application engineering (including algorithm development and benchmarking) even if the degree of realsim is far from being judged as g̈ood lookingb̈y a human observer. The main question to answer

[1] 1http://homepages.inf.ed.ac.uk/rbf/CVonline/Imagedbase.htm
[2] 2http://peipa.essex.ac.uk/index.html

is, to which degree the dataset covers the statistical properties of the real world input data of the image processing algorithm and how do differences between real world and synthetic data affect the results of image processing on the different levels, from single algorithm step result up to overall application ROC. We describe a pipeline how to create synthetic images and refernce data in the following and discuss our approach with the help of two use cases. Output of the content creation pipeline, shown in Fig. 4.1, is the 2D/3D image of the scene and metadata information (ground truth) for every pixel, e.g. the corresponding 3D object, original 3D location in the scene, distance to the camera. The pipeline can be used to render metadata information in higher resolution than the image itself to provide ground truth at subpixel level.

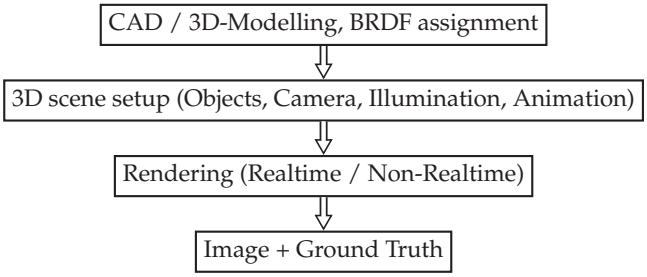


Figure 4.1: Synthetic data creation pipeline.

One of the key advantages of our approach is the possibility, to automatically vary the parameters of the input data creation process on different levels (object position, camera and illumination position, object surface properties), allowing the creation of large datasets without increasing the manual effort. While for single images the time required for capturing and labeling is usually lower than for modeling the virtual scene and rendering, with the increasing number of variations in the input data, the advantage moves towards the synthetic data. For certain scene constellations, e.g. limit samples, sometimes synthetic data is the only way to create these images, because it's too much effort to create those using real world equipment or even required objects are not available.

3.1 Image segmentation for machine vision

This scenario reflects the basic task of image segmentation in machine vision applications by separating one or more solid objects of different shapes and surface properties from each other and the background. As preparation, we measured the BRDFs of the objects using our robot-based goniometer and fitted them to analytical BRDF models [5, 6] Some of the shapes were manufactured based on CAD models, others were measured in the lab with measuring tools and then modeled using CAD. We used the open source software Blender[3] to model the 3D scenes and rendered the images using the open source physically based rendering engine Mitsuba.[4] As a trade-off between rendering quality and speed we choose for bidirectional path tracing, while other algorithms, e.g. Photon Mapping were investigated, too. We added a plugin to the Mitsuba renderer which allowed us to extract the scene content information at rendering time by interfacing with the module calculating the ray-object-intersection. The plugin GroundTruthExtractor exported for every pixel the original 3D position of the object point and the object or object part it belongs to.

Figures 4.2 and 4.3 show examples for scenes with different complexity, allowing for a visual comparison between synthetic image and real camera image. For both environments, a single spotlight light source was used, similar to [1]. We used bidirectional path tracing to render an image of the 3D scene. The real camera hardware was an IDS UI-1460SE color camera, having a resolution of 2048x1536px with a 25mm lens. Ground truth for real images was manually labeled using Adobe Photoshop software.

Figure 4.2: Scene 1: 2 objects, real image (left) vs. synthetic image (right).

[3] www.blender.org
[4] www.mitsuba-renderer.org

Figure 4.3: Scene 2: 5 objects, real image (left) vs. synthetic image (right).

Based on the synthetic images and the ground truth exported during rendering, we evaluated the performance of three different segmentation algorithms: RGB color channel thresholding + morphological operations, Sobel edge detection + morphological operations, Watershed transform.
Figures 4.4 and 4.5 show the results of the RGB-Threshold segmentation algorithm on synthetic data, real image and the resulting error on the real image.

Figure 4.4: Segmentation result on synthetic image (left), real image (middle), difference to ground truth on real image (right).

Figure 4.5: Segmentation result on synthetic image (left), real image (middle), difference to ground truth on real image (right).

Comparing error rates for synthetic and real images shows, that results achieved on synthetic data are close to the results achieved on real images in terms of performance prediction and that the absolute segmentation error rises with the complexity of the scene. We see, that

	Scene 1		Scene 2	
	Synth.	Real	Synth.	Real
RGB-Thresholding	6.9	2.6	17.3	15.2
Sobel Edge Filter	5.9	2.2	28.1	50.8
Watershed Transform	3.1	2.5	71.6	54.1

Table 4.1: Segmentation Error (in percent of foreground pixels).

in case the algorithm has low error compared to the ground truth, it reaches a similar error rate on the real image. In case the error is high on synthetic data, it can be even much higher on real data. Over-Segmentation shown in Fig. 4.5 is identified to be caused by insufficient modeling of the light source characteristics and consequential error in global illumination calculation. Machine vision light sources modeling is currently in our research focus.

3.2 Image segmentation for human-robot-interaction

The second use case discusses the reliable pixelwise segmentation of object classes, in the domain of scene analysis for human-robot interaction (HRI) and human-robot collaboration (HRC). Here a real-world RGB-D sensor is statically mounted on the ceiling and observes the workspace. Environmental objects and human body parts are modeled as different object classes, and training and testing is based on the depth channel of the virtual or respectively the real-world sensor. The prediction model for this use case is a Random Decision Forest (RDF) classifier, where the features of a distinct pixel are the centered depth image patches of fixed size and orientation [7].

Because of the high appearance variation of the objects, especially in case of the human body parts, a large amount of labeled training data must be created for each object class, in order to obtain reliable predictions of the classifier. This is done by automated synthetic depth frame creation in the virtual environment V-REP [8]. Here the static setup of the real-world KINECT sensor is modeled in a virtual scene, and objects are presented to the sensor in random transformations and combinations, which are thought to resemble real-world scenarios. For the human body, we use a parameterized representation of the human body

which allows us to model various body postures during the creation process.

Figure 4.6: *Left*: Synthetic depth data generated with a synthetic KINECT sensor. *Center*: Synthetic depth frame with additive white Gaussian noise. *Right*: Overlay of the object class or respectively body part coloring and the synthetic depth data.

In contrast to the first use case we do not strive for high visual realism in our training data. The used object models are not highly detailed representations of the real-world counterpart. To increase the generalization ability of the classifier we try to produce a large amount of high variant data by using simple models for all object class instances. Especially in case of the human body we only use a coarse approximation which simply consists of a set of spheres in different sizes, aligned along a parameterized skeleton (see Fig. 4.6 *left*). To cope with the high noise levels in the real-world sensor data we apply synthetic noise to the resulting depth frames (see Fig. 4.6 *center*). For the automated creation of the pixelwise ground-truth labeling for each synthetic depth frame, we colored the object class models distinctively (see Fig. 4.6 *right*). The synthetic depth frames are then used in combination with the synthetic RGB frames, in order to sample large numbers of labeled training data for all classes.

Random Decision Forests

We will give a short overview over the principle of RDF training and testing, in order to motivate the different parameters and to describe the weak learner type, which is the basis for the trained decisions in the nodes of the trees. A comprehensive description of RDFs and applications can be found in [9].

A binary Decision Forest F consists of an ensemble of n_t binary Decision Trees $T = \{t_i\}$ with a maximum tree depth $d_{t_{max}}$. A tree t_i has corresponding to its' name a directed binary tree as a graph representation,

with two types of nodes: split nodes which exhibit two child nodes and leaf nodes with no child nodes. Split node represent decisions based on distinct trained feature functions, which are of the same type for all split nodes and trees. Leaf nodes represent the class prediction of a tree. In order to predict a class label of sample s, the sample is routed through the tree according to the decisions of the node feature functions, which process the samples' feature vector $\mathbf{f}(s)$. The leaf node, the sample ends up in, then delivers the prediction for the class label.

When training a tree, a set of training samples with known labels are passed down the tree. In each node the training procedure tries to find the optimal feature function, where optimality considerations are based on quality measures like the entropy.

Training of a forest is done by training the single trees on all training samples, and for testing the empirical class distributions of all trees are used for a forest prediction.

Random Decision Forests are Decision Forests where randomness is injected into the training process, in order to speed up the training and to further the generalization ability and robustness of the classifier. This can be done by randomly choosing subsets of the training samples for the single tree training (bagging), or by randomly choosing fixed sized subsets of feature space dimensions for the decisions in the slit nodes.

In our approach we use both techniques. For the bagging we apply training data sampling with replacement, and for the decisions based on a random feature subspace we use a linear discrimination of 2D subspaces with thresholding of the distance to the linear discrimination border.

Evaluation

For the evaluation of the overall segmentation approach, we use a fixed parameter setup with forest size $n_t = 5$, feature patch size $(w_p, h_p) = (64, 64)$ and maximum tree depth $d_{t_{max}} = 15$. For the randomization in the training process we use 100 threshold and 100 feature function samples in the node optimizations, and bagging with replacement for the tree-wise training data sampling. All training is based on synthetic depth frames with additive white Gaussian noise using a standard deviation of 15 cm. For the performance evaluation we use the Recall and

Precision measure for single object classes and the average as the combined measure for all classes (See [10]).

The numbers presented in Table 4.2 - 4.4, and the prediction results illustrated in Fig.4.7 are based on the same trained decision forest. Here, a total of 5000 synthetic depth frames, generated as described above, were used as a basis for the RDF classifier training. For the training process of each tree, 2000 frames from this data were chosen randomly, and for each frame, 300 pixel positions per object class were chosen uniformly for the extraction of the features patches and ground truth labels. Altogether, this resulted in approximately 2.6×10^6 synthetic labeled training samples per tree, with a training time for the whole forest of approximately 40 min using a PC with Intel i7 CPU and 4 GByte RAM. Calculating the pixelwise predictions for a frame with 640×480 pixels, using the trained forest, takes about 40 ms on this hardware.

When applied to synthetic and real-world testing data, the trained RDF produced similar quantitative and qualitative results for both data types, as presented in Table 4.2 - 4.4 and Fig.4.7 respectively. Overall, the testing of the synthetic data shows better results compared to the real-world data, yet the quantitative measures are not far apart and demonstrate a good overall performance for both types. This indicates, that the training concept based on synthetic data only, using a coarse approximation of the human body in limited postures and transformations is sufficient for the reliable and high-performance segmentation of real-world data, in our application scenario.

Table 4.2: Confusion matrix using synthetic data.

	Bg	He	UB	UA	LA	Ha	L
Bg (Background)	**0.95**	0.00	0.00	0.00	0.00	0.00	0.05
He (Head)	0.00	**0.93**	0.05	0.01	0.01	0.00	0.00
UB (Upper Body)	0.00	0.03	**0.87**	0.08	0.00	0.00	0.02
UA (Upper Arm)	0.00	0.00	0.16	**0.80**	0.04	0.00	0.00
LA (Lower Arm)	0.00	0.00	0.02	0.14	**0.78**	0.06	0.00
Ha (Hand)	0.00	0.00	0.00	0.02	0.23	**0.75**	0.00
L (Legs)	0.00	0.00	0.04	0.00	0.01	0.00	**0.95**

Table 4.3: Confusion matrix using real-world data.

	Bg	He	UB	UA	LA	Ha	L
Bg	**0.95**	0.00	0.00	0.00	0.00	0.00	0.05
He	0.00	**0.84**	0.08	0.02	0.05	0.01	0.00
UB	0.00	0.00	**0.83**	0.15	0.02	0.00	0.00
UA	0.00	0.00	0.19	**0.67**	0.13	0.01	0.00
LA	0.00	0.00	0.00	0.05	**0.77**	0.18	0.00
Ha	0.00	0.00	0.00	0.04	0.15	**0.81**	0.00
L	0.03	0.00	0.04	0.02	0.01	0.03	**0.87**

Table 4.4: Confusion matrix based quality measures.

	Avg	Bg	He	UB	UA	LA	Ha	L
$Recall_{Synth}$	**0.86**	0.95	0.93	0.86	0.79	0.77	0.75	0.94
$Precision_{Synth}$	**0.71**	1.00	0.97	0.79	0.77	0.72	0.63	0.11
$Recall_{Real}$	**0.82**	0.94	0.84	0.83	0.67	0.76	0.80	0.87
$Precision_{Real}$	**0.61**	1.00	0.99	0.70	0.65	0.48	0.46	0.03

4 Conclusion

We presented tools and workflow how to create synthetic image data and image content ground truth information using 3D modeling and computer graphics image synthesis algorithms, which serves as an interim report on our ongoing work in this area. We demonstrated the suitability of synthetic image data and the corresponding ground truth information for selection, training and evaluation of image processing algorithms. This allows for efficient creation of large datasets covering more object and scene variations compared to being created with real cameras and manual labeling. We demonstrated the importance of physically correct modelling and rendering in case the goal is to create an image close to the output of a real vision sensor. In case, process-

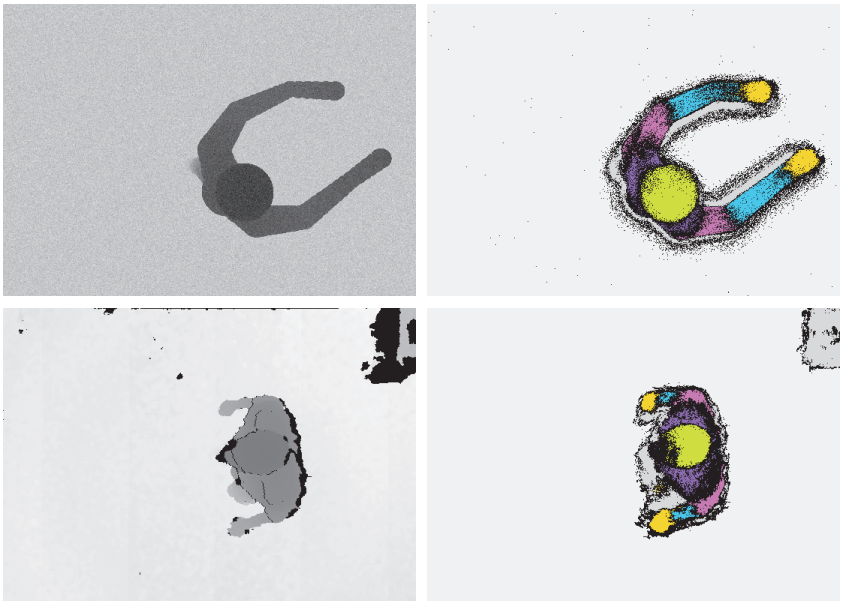


Figure 4.7: Prediction results based on synthetic and real-world data. The first column shows the feature frames based on depth data, the second column shows the prediction results. The first line is based on synthetic testing data, the second line is based on real-world testing data.

ing of the image is not known, this is the goal for the image quality to achieve, which can be reached by using state of the art computer graphics tools and algorithms. We showed, that by knowning the image processing algorithm, the definition of image quality is very different to the expections of a human observer but the benefit for algorithm engineering is high, even if the image is of low visual quality from a human observer's perspective. Despite the highly reduced realism of the object models, the results of our approach show a good and comparable performance for synthetic and real-world testing data in case of the human body classes for depth dasta processing. This motivates the use of synthetic representations in low detail but with high degree of variation in a large training set.

References

1. S. Meister and D. Kondermann, "Real versus realistically rendered scenes for optical flow evaluation," in *Electronic Media Technology (CEMT), 2011 14th ITG Conference on*, 2011, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5936557
2. K. Tarabanis, P. Allen, and R. Tsai, "A survey of sensor planning in computer vision," *IEEE Transactions on Robotics and Automation*, vol. 11, no. 1, pp. 86–104, 1995. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=345940
3. D. Kondermann, "Ground truth design principles," in *VIGTA '13 Proceedings of the International Workshop on Video and Image Ground Truth in Computer Vision Applications*, C. Spampinato, B. Boom, and B. Huet, Eds. ACM Press, 2013, pp. 1–4.
4. S. Meister, "On creating reference data for performance analysis in image processing," Ph.D. dissertation, Universität Heidleberg, Heidelberg, 2014.
5. S. Irgenfried, I. Tchouchenkov, and H. Wörn, "Cadavision: A simulation framework for machine vision prototyping," in *Proceedings of CSSim 2011*, R. Kočí, Ed., 2011, pp. 59–67.
6. R. Gruna and S. Irgenfried, "Reflectance modeling in machine vision: Appliances in image analysis and synthesis," in *Machine Vision - Applications and Systems*, F. Solari, M. Chessa, and S. P. Sabatini, Eds., 2012, pp. 227–246. [Online]. Available: http://dx.doi.org/10.5772/26554
7. J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011. [Online]. Available: http://research.microsoft.com/pubs/145347/BodyPartRecognition.pdf
8. Coppelia Robotics, "v-rep: virtual robot experimentation platform," 31.05.2014. [Online]. Available: http://coppeliarobotics.com/
9. A. Criminisi and J. Shotton, *Decision Forests for Computer Vision and Medical Image Analysis*. Springer Publishing Company, Incorporated, 2013.
10. M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, Jul. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.ipm.2009.03.002

Synthetic image acquisition and procedural modeling for automated optical inspection (AOI) systems

Max-Gerd Retzlaff, Josua Stabenow and Carsten Dachsbacher

Karlsruhe Institute of Technology (KIT)
Institute for Visualization and Data Analysis, Computer Graphics Group
Am Fasanengarten 5, 76131 Karlsruhe

Abstract When designing or improving systems for automated optical inspection (AOI), computer graphics methods can be used to quickly create large virtual sets of samples of test objects and to simulate image acquisition setups. We demonstrate this approach for shards of glass as sorting glass is one practical application for AOI. We use procedural modeling techniques to generate virtual objects with varying appearance and properties, mimicing real objects and sample sets. A physics engine is used to simulate the placement of virtual objects, and physically-based rendering techniques are used to create synthetic images. These are used as input to an AOI system instead of physically acquired images. This enables the development, optimization, and evaluation of the image processing and classification steps of an AOI system independent of a physical realization.

1 Introduction

An important problem in the development, optimization, and evaluation of an automated optical inspection (AOI) system consists in the availability of annotated image data at an early stage. Usually, a configuration or lab setting of the AOI system has to be built, maintained, and operated before any image acquisition can take place. Often it would be helpful to have images at hand long before an acquisition system is developed and operational. But even if a system is already present the availability of testing objects may be another concern. The desired objects may simply not be accessible or yet available. In other cases it may

be decaying objects like fruit, or a group of objects with a large variety as it is with most natural objects. In such cases it may not be feasible to have a sufficiently large sample of test objects that comprises the variety range of the testing object.

Our solution for this problem is twofold. We propose to develop procedural models that describe the instances of a selected sample of a specific object class. These models can be used to generate new instances that could be part of the original sample. Physically-based rendering techniques can then be used to synthesize realistic images of these generated, virtual objects. These images can be fed into the image processing and classification process of a physical or simulated AOI system.

Computer graphics methods can quickly generate synthetic images in large variety, simulating different surrounding conditions and disturbing factors. At the same time full annotation information is always available.

2 Diversity of possible objects of interest

The range of objects for which methods of (automated) optical inspection are used is utterly diverse. Natural objects with a great variety such as fruit can be classified and sorted, but AOI is also applied to artificial objects such as certain cast machine parts in order to identify defects.

A procedural model that comprises even only these two major groups of objects would have to be exceedingly broad and therefore more specialized models have to be used. In this paper we present an application of our proposed solution process in the field of sorting and classification of bulk goods in the form of shards of glass waste (cf. figure 5.1).

3 Outline of the methodological procedure

First of all, a procedural model has to be attained that is able to produce instances of an object class with a potentially great variety. This is accomplished by these steps:

1. assembly of a representative ensemble of objects,
2. measurement of relevant properties (in the case of glass shards, e. g., dimensions, geometry, and light extinction coefficients),

Figure 5.1: Photo of real glass shards obtained by a line scan camera of an AOI image acquisition system (courtesy of Fraunhofer IOSB).

3. measurement of defects and surface details, and
4. design of the procedural model.

Images of virtual shard scenes are attained in this way:

1. generation of virtual objects using the procedural model,
2. placement and distribution of the objects in realistic scenes, and
3. image synthesis (or rendering) of the scenes.

The produced images can be displayed, stored, or directly fed into a physical or simulated AOI system.

We evaluate our approach in this manner:

1. acquisition of ground truth,
2. comparison of our synthetic images to the ground truth, and
3. use of the synthetic images as input to an AOI system instead of physically acquired images.

4 A procedural model for glass shards

In order to develop a procedural model that is capable of generating realistic synthetic glass shards, we have measured a sample of one thousand glass shards with a number of methods.

On one hand, the data is used to directly improve the simulated glass shards. For example, we can obtain extinction coefficients by measuring shards and use these coefficients in the simulation. One the other hand, we can use the obtained data as ground truth to assess the quality of the simulated shards using defined metrics as, for example, the distribution of Fourier descriptors, as outlined in section 6.2.

4.1 Assembly of a glass shard ensemble

It is important to use shards that have been subjected to the same process that usually ends in a sorting system, and not to produce shard samples by just breaking glass in a different process as a) the broken down shards are repeatedly relocated in large quantities and this leads to shards that have lost all sharp edges. And b) the whole process of breaking to pieces in recycling containers, the transportation, shredding, and sieving using meshes of different grid sizes leads to rounded shards in their overall appearance and to characteristic size distributions. At the same time, the shards get small point-shaped damages on their surfaces.

We received a batch of glass waste from a producer of glass sorting machines and project partner of Fraunhofer IOSB. We randomly selected from this sample in order to obtain an ensemble of one thousand shards.

4.2 Measurement of real glass shards

We conducted the following measurements:

1. high-dynamic range (HDR) images using exposure bracketing of image series using a high-quality camera and a telecentric lens,
2. systematical measurement of the thickness of the individual shards with a sliding gauge, and
3. determination of the extinction coefficients.

In order to obtain a ground truth in addition to the just mentioned measurements, we captured reference images using a lab setup of a physical image acquisition system that is used in glass sorting systems, as detailed in section 6.1, Acquisition of ground truth.

Telecentric HDR Images As in image acquisition systems of sorting machines for glass waste, the transmitted light of the glass shards is captured while the shards are located in front of a diffuse light source. In the acquisition setup an LED light box is used as light source. The deployed telecentric lens, an Opto Engineering *TC 23 096*, produces an orthographic view of the shards that can be used to measure the circumference and area of the shards easily. To capture the images we use the 5 megapixel high-quality microscope camera Leica *DFC425*. This camera possesses a linear sensitivity that has been verified at Fraunhofer IOSB as part of an earlier project.

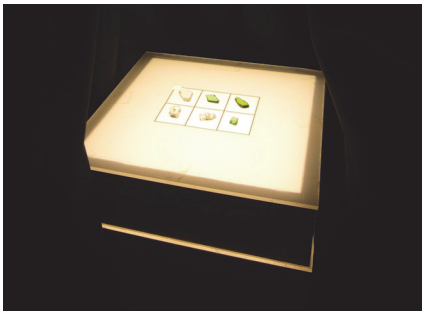


Figure 5.2: Taking photos of six shards at a time on an LED light box.

The glass shards are photographed six at a time and positioned in a stencilled frame that features marks at 1 mm intervals (cf. figure 5.2). Images are taken in standard exposure steps with eleven exposure times of 1 s, 0.5 s, ..., 1 ms. Inhomogeneities of the illumination are levelled out using averaged reference images, and the resulting images are fused to an HDR image while over- and underexposed pixel values are masked. As the original images are preserved, the image fusion could be replaced by a more elaborate image fusion should the necessity arise. As of now, we do not correct chromatic aberration of the acquired images, but a series of test charts to determine aberration has been recorded to allow for a later correction. Each shard set is photographed two times as most shards are rather planar and as such exhibit two obvious orientations.

Measurement of the shard thicknesses The thickness of each glass shard of the sample has been measured using a sliding gauge. These single point measurements have an accuracy of 0.1 mm or better. At the same time, a reference point is placed on the acquired two images (front and backside) of the particular shard.

In case of shards that exhibit an irregular thickness we take several measurements to approximate the marked reference point and calculate the sample standard deviation taking into account a correction factor assuming a normal distribution to obtain an unbiased estimation of the calculated standard deviation.

Determination of the extinction coefficients Out-scattering and absorption of light are not determined separately but combined as the extinction coefficient. We experimented with an acquisition setup to directly measure the light extinction of a white light source using the high-resolution spectrometer Ocean Optics *HR2000+*. However, we opted to determine the coefficients using the high quality HDR images acquired by the telecentric setup in connection with the measured thicknesses, as both measurements are already available.

We compute the mean color of the reference point neighborhood (diameter 64 pixels) as well as the mean color of the unobstructed light of the light box (taken from the same image). Then we calculate the extinction coefficients using the Beer-Lambert law with the previously measured thicknesses and the light intensities obtained by taking the computed mean colors relative to their exposure times. The standard deviations of the mean colors and the thickness measurement are propagated according to the Gaussian error propagation law.

Reconstruction of the shards' 3D geometry To complete our acquisition of glass shards, we plan to measure height profiles of a (small) collection of shards using the 3D line scan measurement device SICK *Ruler-E2122*. By global registration of the height profiles we want to reconstruct the 3D shapes.

4.3 Measurement of defects and surface details

The measurement and description of defects and surface details that the objects of interest can exhibit is an important and non-trivial task separate of the measurement of the relevant overall properties. Surface details often arise as characteristic consequence of the production or handling processes. Cast machine parts, for example, have a characteristic surface and specific types of defects, as have natural objects such as fruit.

Of course, glass waste exhibits specific kinds of defects as well. We inspected the breaking edges and surfaces of a small selection of glass shards by measuring the surface topography based on depth from focus using the 3D reconstructing system Alicona *InfiniteFocus* and by confocal microscopy as well as white light interferometry, both supported by the Leica *DCM 3D*. We noted that the spots on the shard surfaces are actually not small soilings but tiny surface damages and holes.

To synthesize shards of glass waste that are as realistic as possible, these surface defects have to be measured and described so that the resulting model can reproduce them, which we plan for future work.

4.4 Design of the procedural model

As basis for the procedural model we use a simplified version of an algorithm described by Martinet et al. [1]. It consists of a procedural method for modeling cracks and fractures in solid materials. Fragments are generated by recursively splitting the initial object using a *carving volume* such as a sphere or an ellipsoid. The main difference is that the algorithm of Martinet et al. does not only describe the generation of separate shards. Instead a procedural model is presented that generates crack patterns and fractures *on an object*, i. e., a whole initial object is divided into interlocking fragments.

In our method we instead concentrate on generating plausible shards while it is not important for us that they can be combined to recreate the initial object. To actually compute a breakup of several initial objects and randomly select from all generated shards would introduce a huge overhead, as a large number of fragments that are going to be discarded is generated. In addition to this, a smoothing of the shards is permitted, or rather desired. The reasoning is that the shards of a random sample

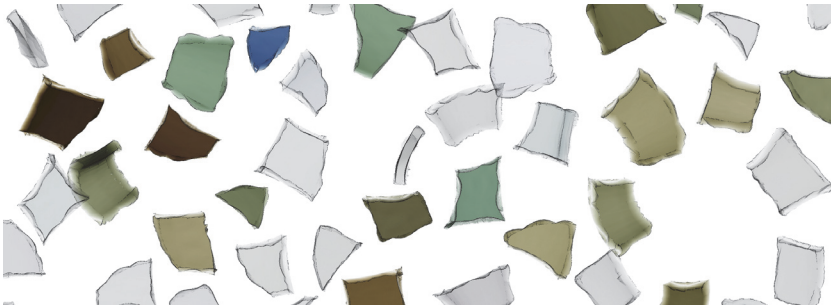


Figure 5.3: Synthetic image of glass shards, generated and rendered in real-time by our implementation.

of glass waste that was repeatedly relocated are quite well mixed, and basically consist of random shards from a large number of initial objects. At the same time, the transportation and relocation process leads to rounded edges of the glass shards.

Of course, there are methods to generate shards and crack patters in a non-phenomenological way, like, for example, finite element methods to determine stress patterns on object surfaces, as described, e. g., by Iben and O'Brien [2].

5 Generation of images of virtual shard scenes

The task of generating synthetic shard images is split into three steps: the virtual objects have to be generated, the objects have to be placed and distributed in order to obtain realistic scenes, and the scenes have to be synthesized (or rendered) as images.[1]

5.1 Generation of virtual glass shards

We use our implementation of the procedural model described in section 4.4 to generate virtual glass shards. These shards are in line with the selected ensemble of physical shards that were selected in order to obtain the procedural model. Of course, ground-truth measurements of

[1] An earlier version of our method is documented in [3].

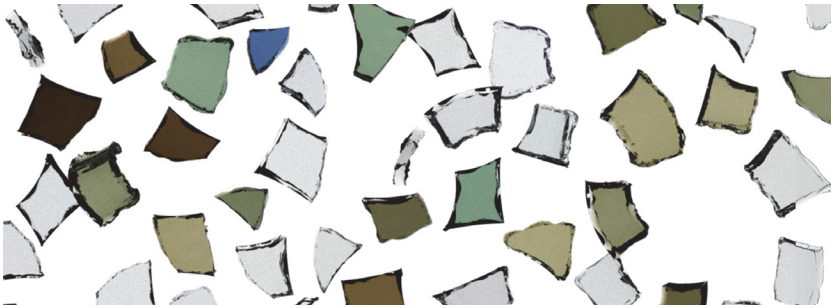


Figure 5.4: Synthetic image of glass shards, generated by our implementation and rendered by a physically-based rendering method.

the shape and details of real shards are necessary to make sure of this, as described in section 6, Evaluation.

5.2 Compilation of realistic glass scenes

We use the physics engine JBullet [4] to achieve realistic placements of the glass shards. JBullet implements collision detection and a rigid body simulation, hence the virtual shards influence each other as in reality, i. e., they can bounce off one another or lie on top of each other in realistic way. We can also simulate the fall by gravity and render images reproducing the effects that physical cameras capture in such circumstances.

In summary, we assert that our implementation can simulate the movement of glass shards during the sorting process as it happens in a physical sorting system. The virtual shards are placed and distributed in a way that realistic scenes are accomplished.

5.3 Image synthesis

Our implementation includes a real-time GPU-based rendering of the shard distributions. A screenshot is shown in figure 5.3. The light extinction within the glass volume is computed using the determined extinction coefficients (cf. section 4.2).

Hyperspectral images The idea to select certain additional wavelengths in addition to the usual RGB or even grayscale representations in the field of AOI systems for sorting of bulk goods is presented and discussed, for example, in [5].

Our method is not limited to the generation of RGB images. The implementation can easily be extended to other sampling points of the light spectrum in addition to the standardized CIE RGB frequency ranges, for example in the UV or NIR range.

External renderers The generation of virtual shards and realistic scenes thereof is not bound to be used in connection with our real-time rendering implementation. The shard distributions can be exported as scene files suitable for external renderers. For figures 5.4 the Mitsuba Renderer [6] was used, a state of the art rendering system capable of, amongst others, spectral and volume rendering with participating media.

6 Evaluation

We evaluate our method by comparing the synthetic images to a measured ground truth and by using the images instead of physically acquired images. This work has only started and we describe our current and intended activities.

6.1 Acquisition of ground truth

The final aim of our activities is to synthesize images of glass shards comparable to the images of real glass shards taken by a physical image acquisition system. We obtained reference images with a lab setup of the *HR Fine* image acquisition system by Fraunhofer IOSB that is part of glass sorting systems (shown in figure 5.5) to capture images of our ensemble of one thousand glass shards.

6.2 Comparison to ground truth

A first step is to compare the synthetic images of individual shards to the telecentric HDR images. Next, we compare the images of virtual

Figure 5.5: Image capturing using a lab setup of the *HR Fine* image acquisition system at Fraunhofer IOSB.

shard scenes to our captured images of the physical image acquisition system.

It is important to determine and develop suitable metrics for the evaluation. We experimented with complex Fourier descriptors as a means to encode the shape of the shards' two dimensional contour curves, and intent to use a metric for shape differences to measure the distances between Fourier descriptors.

6.3 Use of synthetic images for an AOI system

Our goal is to use synthetic images as input to an AOI system instead of physically acquired images. Although we could already feed the synthetic images into an existing lab setup of an image processing and classification system, this has not been tried yet but is planned for the near future.

Recently, a first application of our methods emerged in the development of new classification systems based on machine learning as full annotation information is always available for the synthetic images.

7 Conclusion

As in many other fields of application for image processing, synthetic images created using computer graphics methods (and in particular procedural modeling) show great potential for the design and improvement of AOI systems. The presented method lays the foundation for the development, optimization, and evaluation of the image processing and classification steps of an AOI system without a need for a physical realization, real test objects, or physically acquired images in general.

The systems can be trained and evaluated with a greater number and variety of glass shards compared to conventional lab settings and with diversified waste distributions. In addition to this, synthetic image acquisition can easily simulate different surrounding conditions and disturbing factors as, for example, changing lighting conditions, soiling, turbidity, or scratches on parts of the optical system and image sensor.

Automated optimization is made easier to implement than with physical image acquisition because synthetic images are always fully annotated without additional effort.

Acknowledgments

The first author is funded by the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB). We thank Prof. Beyerer for his valuable input. The Fraunhofer IOSB provided access to its extensive collection of microscopic equipment in the IOSB MicroLab, while the IOSB department SPR made its *HR Fine* image acquisition system and a high-resolution spectrometer available. Binder+Co AG supplied us with a batch of shards of glass waste that we used as sample.

References

1. A. Martinet, E. Galin, B. Desbenoit, and S. Hakkouche, "Procedural modeling of cracks and fractures," in *Proceedings of the Shape Modelling International (Short Paper)*, Washington, DC, USA, 2004, pp. 346–349.
2. H. N. Iben and J. F. O'Brien, "Generating surface crack patterns," in *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, Vienna, Austria, Sept. 2006, pp. 177–185.
3. J. Stabenow, *Procedural Modeling of Glass Fragments* Karlsruhe: Karlsruhe Institute of Technology, 2014, bachelor thesis.
4. M. Dvorak, "JBullet – Java port of Bullet Physics Library," 2008–2010, http://jbullet.advel.cz.
5. M. Michelsburg, R. Gruna, K.-U. Vieth, and P. L. Fernando, "Spektrale Bandselektion beim Entwurf automatischer Sortieranlagen," in *Forum Bildverareitung*. Karlsruhe: KIT Scientific Publishing, 2010, pp. 131–142.
6. W. Jakob, "Mitsuba – Physically Based Renderer," 2010, http://www.mitsuba-renderer.org.

Verdeckungs- und affin-invarianter Regionendetektor basierend auf Farb- und Frequenzinformation

Pilar Hernández Mesa, Ron Heiman und Fernando Puente León

Institut für Industrielle Informationstechnik,
Karlsruher Institut für Technologie,
Hertzstraße 16, 76187 Karlsruhe

Zusammenfassung In diesem Beitrag wird ein neuartiger verdeckungs- und affin-invarianter Regionendetektor vorgestellt. Dieser ist in der Lage, zusammengehörende Musterkombinationen – wie Schachbrett- und Zebramuster – in Farbbildern zu identifizieren. Der Detektor besteht aus zwei Stufen. Die erste orientiert sich am MSCR-Detektor (*maximally stable colour regions*) und erweitert diesen, indem überschneidungsfreie, zusammenhängende Farbbereiche identifiziert werden. In der zweiten Stufe werden die zusammengehörenden Musterkombinationen mit Hilfe einer Orts-Frequenz-Analyse und der Verwendung von Graphendarstellungen erkannt.

1 Einleitung

Die Erkennung zusammengehörender Bildbereiche stellt einen wichtigen Schritt innerhalb der Bildverarbeitung dar. Die Bildbereiche sollen möglichst unabhängig von Verdeckungen, affinen Transformationen (Verschiebungen, Drehungen, Stauchungen und Verzerrungen der Szene) sowie Kompressionsungenauigkeiten des Bildes detektiert werden. Anwendungsgebiete finden sich beispielsweise in der Robotik sowie in der inhaltsbasierten Bildsuche.

In der Literatur gibt es zahlreiche Vorschläge zur Gewinnung von solchen Regionen, wie beispielsweise den MSER-Detektor (*maximally stable extremal regions*) oder den SURF-Detektor (*speeded-up robust features*) [1, 2]. Die heutzutage am häufigsten verwendeten affin-invarianten

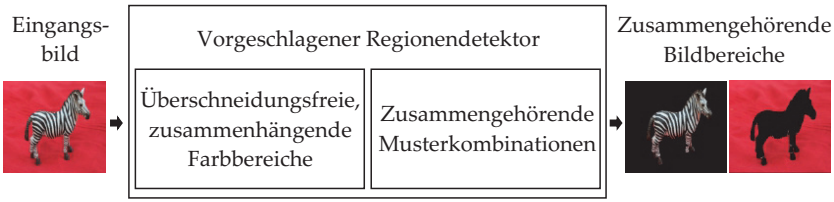


Abbildung 6.1: Ablauf des vorgestellten Regionendetektors.

Detektoren untersuchen jedoch lediglich Grauwertbilder [3]. Um zusammengehörende Farbbereiche innerhalb von Bildern zu detektieren, wird in [4] das MSCR-Verfahren vorgeschlagen. Die Erkennung zusammengehörender Bildbereiche, welche ein Muster aus mehreren Farben darstellen, ist aber mit dem MSCR-Detektor weiterhin nicht möglich. Periodische Muster, wie zum Beispiel Schachbrett- und Zebramuster, werden nicht als zusammengehörende Bildbereiche erkannt. In dieser Arbeit wird ein Verfahren vorgestellt, um solche Regionen unabhängig von Verdeckungen und affinen Transformationen zu detektieren. Eine Übersicht über das Verfahren ist in Abbildung 6.1 zu sehen. Hierzu werden zunächst die Verfahren zur Bestimmung überschneidungsfreier, zusammenhängender Farbbereiche (Abschnitt 2) und darauffolgend, die Bestimmung zusammengehörenden Musterkombinationen (Abschnitt 3) erläutert. In Abschnitt 4 werden die Ergebnisse dargestellt und diskutiert, um in Abschnitt 5 den Beitrag zusammenzufassen.

2 Überschneidungsfreie, zusammenhängende Farbbereiche

Überschneidungsfreie, zusammenhängende Farbbereiche sind der Ausgangspunkt für den zweiten Teil des Detektors, wo sie zu Musterkombinationen zusammengefasst werden.

2.1 Stabile Farbbereiche

Die hier verwendete Methodik lehnt sich an den MSCR-Detektor [4] an, welcher eine Erweiterung des MSER-Detektors [1] darstellt.

Der MSER-Detektor untersucht Grauwertbilder, indem iterativ alle möglichen Intensitätswerte mittels eines Schwellwertvektors $\boldsymbol{d}_{\text{thr}} = [0, \ldots, 255]$ durchgegangen werden. In jedem Iterationsschritt i werden die Pixel betrachtet, deren Intensitätswerte kleiner sind als der Schwellwert $\boldsymbol{d}_{\text{thr}}(i)$. Diese werden zu zusammenhängenden Regionen zusammengefasst. Werden in einer Iteration zwei unterschiedliche Regionen zusammengefasst, so wird die kleinere von der größeren übernommen. Als stabile Regionen werden dann diejenigen bezeichnet, deren Fläche sich über verschiedene Schwellwerte nur geringfügig ändert.

Beim MSCR-Detektor [4] werden Farbbilder betrachtet. In einem ersten Schritt wird die Farbdifferenz zwischen benachbarten Pixel berechnet, wobei die Verwendung des Chi-Quadrat-Abstands in [4] empfohlen wird. Analog zum MSER-Detektor werden anschließend die Differenzen iterativ mit einem Schwellwert verglichen. Dieser wird in [4] mit Hilfe des erwarteten Verlaufs der Wahrscheinlichkeiten zwischen Nachbarpixeln hergeleitet. Stabile Regionen werden durch die Berücksichtigung der Änderung der Fläche und ihrer Schwellwerte ausgewählt.

In Anlehnung an den MSCR-Detektor wird hier zuerst das Differenzbild bestimmt. Dieses ergibt sich aus dem Mittelwert der euklidischen Abstände der Farbwerte zum nächsten obigen und rechten Nachbarn. Anschließend wird das Differenzbild mit einem Schwellwertvektor verglichen, der empirisch aus der inversen Verteilungsfunktion der Differenzbilder von Testbildern bestimmt worden ist. Die Festlegung von stabilen Regionen verläuft weiterhin wie beim MSCR-Detektor. Um Rauschen zu unterdrücken, werden zwei Differenzbilder w_1, w_2 bestimmt. Für beide wird das ursprüngliche Bild vorab mit einem Gauß-Filter geglättet. Bei w_2 werden aber an den Kanten die ursprünglichen Farbwerte des Bildes behalten. Die stabilen Regionen beider Differenzbilder werden zusammengefasst. Pixel in stabilen Regionen aus w_2 werden aus den Regionen von w_1 entfernt.

2.2 Überschneidungsfreie Farbbereiche

Ausgangspunkt sind die stabilen Farbbereiche aus dem vorherigen Abschnitt, welche nicht überschneidungsfrei sein müssen. Jede Region wird als ein einzelner Knoten v_m betrachtet und das sich daraus zusammengesetzte Bild durch einen Graphen beschrieben. Knoten von Regionen, die sich überschneiden, werden nur dann durch eine Kan-

te verbunden, wenn es keine weitere Region gibt, die eine Teilmenge der größeren ist und welche die kleine enthält:

$$(v_a, v_b)\,, \text{ falls } v_a \supset v_b \text{ und } \nexists v_c : v_a \supset v_c \supset v_b\,.$$

Hiermit entstehen einzelne Knoten für die Farbregionen, die keine anderen Regionen enthalten, und Bäume für die sich überschneidenden Regionen (siehe Abb. 6.2 a, das Verfahren wird an dem hervorgehobenen Teil des Baums veranschaulicht). Jeder Knoten v_m, der Kinderknoten v_s^{Kind} enthält, wird weiter untersucht. Die Menge der Pixel $\mathcal{P}_m^{\text{R1}}$, welche in der Region des Knotens v_m enthalten sind, aber nicht in deren Kinderknoten v_s^{Kind}, wird vorgemerkt:

$$\mathcal{P}_m^{\text{R1}} = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \middle| \begin{pmatrix} x \\ y \end{pmatrix} \in v_m \text{ und } \begin{pmatrix} x \\ y \end{pmatrix} \notin \bigcup_{s \in \mathbb{N}} v_s^{\text{Kind}} \text{ mit } v_m \supset v_s^{\text{Kind}} \right\}.$$

Für diese Pixel werden die Bereiche bestimmt, die örtlich zusammenhängend sind. Alle Bereiche mit einer Mindestfläche $\mathcal{P}_m^{\text{R1},g}$ $(1 \leq g \leq G)$ werden als neue Kinderknoten von v_m hinzugefügt. Sie stellen Blattknoten im Baum dar. Die übrigen Pixel werden als Restpixel $\mathcal{P}_m^{\text{R2}}$ der Region vorgemerkt (siehe Abb. 6.2 b, wo die Knoten v_4 und v_2 weiter zerlegt wurden).

Im Folgenden wird iterativ ausgehend von den Knoten mit der kleinsten Fläche bis zur größten Fläche durchgegangen (von den Blättern bis zur Wurzel). Betrachtet werden nur diejenigen Knoten, die keine Einzel- und Blattknoten sind. Enthält der aktuelle Knoten nur einen Kindknoten, so wird dieser Kindknoten verworfen (siehe Abb. 6.2 c, wo v_1 verworfen wird). Enthält dieser jedoch mehrere Kinderknoten, so werden weitere Untersuchungen vorgenommen. Zuerst wird ein Hilfsgraph erstellt, in dem nur die Kinderknoten dargestellt werden. Diese werden mit einer Kante verbunden, wenn deren mittlere Farbdifferenz hinreichend klein ist (siehe Abb. 6.2 d). Wird der Hilfsgraph zusammenhängend, so werden die Kinderknoten verworfen. Im Falle keines zusammenhängenden Hilfsgraphen werden die einzelnen Knoten als Knoten in den ursprünglichen Graphen übernommen. Knoten, die im Hilfsgraphen mit einer Kante verbunden wurden, werden zusammen mit den Restpixeln Γ_m^{R2} nach örtlich zusammenhängenden Bereichen untersucht und diese als Knoten in den ürsprunglichen Baum

übernommen. Die verbliebenen Kinder und neuen Knoten übernehmen die übergeordneten Regionen ihrer ehemaligen Vorgänger (siehe Abb. 6.2 e, die Knoten v_2 und v_8 nehmen als übergeordnete Region, die von deren ehemaligen Vorgänger v_4).

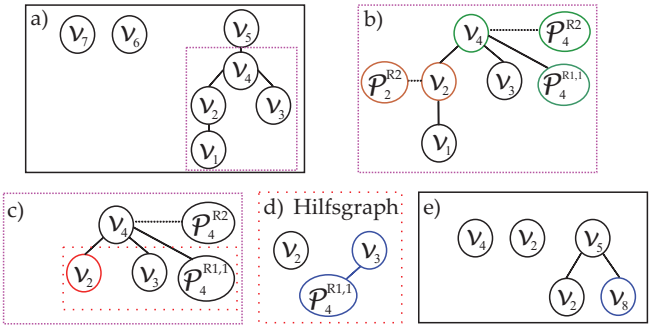


Abbildung 6.2: Beispiel zur Erstellung überschneidungsfreier Farbbereiche.

Nach der fertigen Bearbeitung aller Bäume entstehen K überschneidungsfreie, zusammengehörende Farbbereiche n_k $(1 \leq k \leq K)$.

3 Zusammengehörende Musterkombinationen

Ziel des zweiten Teils des Algorithmus ist die Erkennung zusammengehörender Musterkombinationen.

3.1 Orts-Frequenz-Analyse

Für viele Signale ist eine Untersuchung ihres Orts-Frequenz-Verhaltens sinnvoll. Die Wavelet-Transformation bietet hierzu eine Möglichkeit und kann effizient mit Hilfe von Filterbänken realisiert werden. Das Signal wird durch verschiedene Stufen n jeweils hochpass- und tiefpassgefiltert (Detail- und Approximationskoeffizienten), was einer Projektion in Unterräume entspricht [5, 6]. Um die Wavelet-Transformation auf zweidimensionale Signale anzuwenden, wird sie auf die Zeilen und anschließend auf die Spalten des Signals angewendet [6]. Aufgrund der Zweidimensionalität des Signals erhält man pro Stufe eine Matrix mit

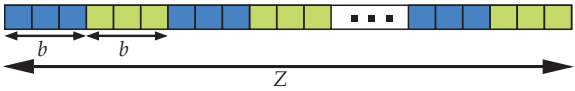


Abbildung 6.3: Periodische Strukturen, die erkannt werden sollen.

den Approximationskoeffizienten sowie drei Matrizen, die die Detailkoeffizienten enthalten, die zur Untersuchung von horizontalen, vertikalen und diagonalen Strukturen genutzt werden können. Innerhalb dieser Arbeit wird die Anzahl der Stufen der Wavelet-Transformation bestimmt, indem zuerst die Strukturen (b, b) (b Pixel einer Farbe gefolgt von b Pixel einer andere Farbe) festgelegt werden, die mit einer gesamten Länge Z gefunden werden sollen (siehe Abb. 6.3). Mit deren Hilfe lassen sich für die Breite L und die Höhe H des Bildes die normierten Frequenzen bestimmen, die in der höchsten Stufe (n_{Hor}, n_{Ver}) der Wavelet-Transformation durch die Hochpassfilterung durchgelassen werden sollen:

$$f_{\text{Norm,Hor}} = \frac{Z}{b} \cdot \frac{1}{L}\,, \qquad f_{\text{Norm,Ver}} = \frac{Z}{b} \cdot \frac{1}{H}\,. \tag{6.1}$$

Für die horizontale und vertikale Analyse des Bildes werden die Stufenanzahlen der Wavelet-Transformation n_{Hor} und n_{Ver} gewählt, bei denen zum ersten Mal die Frequenzen aus (6.1) von der Hochpassfilterung durchgelassen werden. Der kleinere der beiden Werte bestimmt die Anzahl der Stufen für diagonale Strukturen n_{Diag}.

Für alle Farbkanäle des Bildes werden zunächst die Detailmatrizen über alle Stufen und alle drei Richtungen bestimmt. Diese Unterprojektionen werden jeweils getrennt auf den örtlichen Raum des Bildes zurückprojiziert. Pro untersuchter Strukturrichtung (horizontal, vertikal und diagonal) wird schließlich eine Matrix der Größe des Bildes bestimmt, welche ein Maß für die Wahrscheinlichkeiten liefert, dass jedes Pixel Teil einer Musterkombination ist. Hierfür wird für jedes Pixel pro Richtung der maximale Wert über alle Farbkanäle sowie über die zurückprojizierten Detailkoeffizienten bestimmt. Schließlich werden diese drei Matrizen mit einem Gauß-Filter geglättet.

3.2 Vorbereitungen der Farbregionen

Für die gefundenen zusammenhängenden, überschneidungsfreien Farbregionen n_k aus Abschnitt 2.2 werden folgende Eigenschaften definiert.

Die mittlere Farbe und Ausdehnung wird für jede Region festgelegt. Weiterhin wird jeder Farbregion ein Frequenzwert zugeordnet. Dieser entsteht, indem für jedes Pixel der Region dessen maximaler Frequenzwert über alle drei Richtungen bestimmt wird. Das p-Quantil dieser Werte ergibt die zugeordnete Frequenz.

Für jede Farbregion n_k werden deren Nachbarregionen bestimmt. Hierfür werden zuerst die Menge der Pixel auf deren Rand ∂n_k bestimmt. Als Nachbarschaft $\mathcal{N}(n_k)$ werden vorerst diejenigen Regionen definiert, welche mindestens eine der folgenden Bedingungen erfüllen ($n_w \in \mathcal{N}(n_k)$ mit $w \neq k$):

1. $\|\boldsymbol{x}_k - \boldsymbol{x}_w\| \rightarrow \min$, $\boldsymbol{x}_k \in \partial n_w$ und $\boldsymbol{x}_w \in \partial n_w$
2. $\|\boldsymbol{x}_k - \boldsymbol{x}_w\| \leq \delta_{\text{thr}}$, $\boldsymbol{x}_k \in \partial n_k$ und $\boldsymbol{x}_w \in \partial n_w$
3. $n_w \in \mathcal{N}(n_k) \rightarrow n_k \in \mathcal{N}(n_w)$

Als Grad für die Nachbarschaft $S^{\mathcal{N}}_{n_k}(n_w)$ der Region n_w zu n_k wird die Anzahl der Pixel aus ∂n_k bestimmt, welche zu n_w den minimalen euklidischen Abstand haben. Von den vorgemerkten Nachbarregionen werden für jede Farbregion diejenigen behalten, deren räumlicher euklidischer Abstand kleiner als der Mittelwert der mittleren Ausdehnung der betrachteten Farbregionen ist. Ist der Wert größer, so werden die Regionen nur dann als Nachbar behalten, falls die mittlere Farbe der Pixel zwischen den beiden Regionen, die keiner Farbregion zugeordnet wurde, entweder nah genug an der mittleren Farbe einer der beide Regionen liegen oder in den drei Farbkanälen zwischen dem maximalen und dem minimalen Wert der Regionen liegen.

In einem neuen Graphen $\mathcal{G}$ werden die Farbgruppen als Knoten dargestellt und benachbarte Farbgruppen durch eine Kante verbunden.

Durch den Menschen als zusammenhängend erkannte Regionen könnten beispielsweise wegen Kompressionsungenauigkeiten mit der Methodik aus Kapitel 2 nicht erkannt werden. Um solche Regionen zu berücksichtigen, werden hochfrequente, benachbarte Regionen mit ähnlicher Farbe genauer betrachtet. Besitzen solche Regionen keine

gemeinsamen Nachbarn oder ist der Grad der Nachbarschaft größer als zu ihren gemeinsamen Nachbarn, so wird nach einer möglichen zusätzlichen Region gesucht. Diese wird als neue Region n_k hinzugefügt, falls es zwischen den beiden Regionen genügend Pixel gibt, die keiner Farbregion zugeordnet wurden und bestimmte Voraussetzungen in Farbe und Frequenz erfüllen. Der Frequenzwert der neuen Region ergibt sich allerdings aus dem mittleren Wert, der aus den Maxima der Frequenzen der drei Richtungen pro Pixel bestimmt wird.

Weiterhin werden benachbarte Farbregionen mit ähnlichen Farben als ein Farbcluster Ξ_v $(1 \leq v \leq V)$ vermerkt.

Schließlich werden Farbregionen und Farbcluster, die hauptsächlich im Rahmen des Bildes liegen, als Hintergrundbereiche bemerkt.

3.3 Mustererkennung

Ausgehend von den Farbregionen sowie der Orts-Frequenz-Analyse wird im Folgenden nach Musterkombinationen gesucht (siehe Abb. 6.4). Bei der Initialisierung (Abb. 6.4 a) wird die Iterationsliste der Farbbereiche bestimmt, die Teil eines Musters sein könnten. Pro Farbbereich werden die Pixel berücksichtigt, die nicht Teil des Randes sind. Für jede Region wird der maximal auftretende Wert der Frequenz sowie dessen dazugehörige Richtung der Struktur bestimmt. Die Liste entsteht

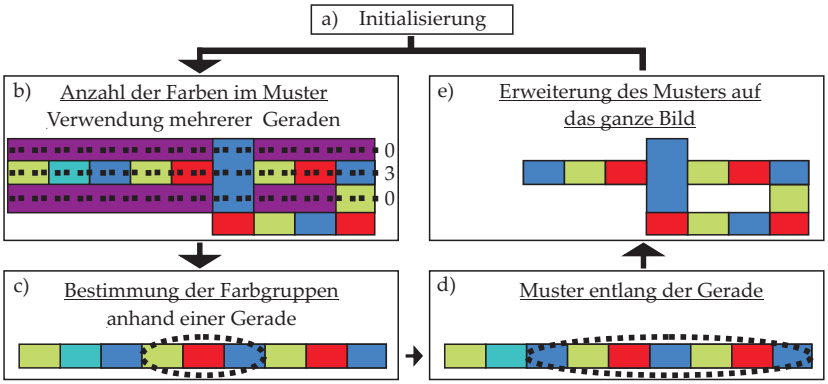


Abbildung 6.4: Ablauf des vorgestellten Detektors.

aus der absteigenden Sortierung der Farbbereiche nach der Maximalfrequenz, wenn diese eine Mindestgröße erreichen. Iterativ wird jeder Bereich bearbeitet, der nicht Teil der Menge $\mathcal{R}$ (bearbeitete Regionen) ist, die leer initialisiert wird.

Für den betrachteten Farbbereich n_k wird die Anzahl der auftretenden Farben im Muster festgelegt (Abb. 6.4 b). Verwendet werden Geraden, die parallel zu der Struktur des Farbbereichs verlaufen und diesen enthalten. Im Falle einer diagonalen Struktur werden zwei Geradengruppen (Haupt- und Nebendiagonale) festgelegt. Ausgehend von der betrachteten Farbregion innerhalb der Geraden werden die Farbabstände zu den anderen Regionen entlang der Geraden bestimmt. Die nächste Region mit einer hinreichend kleinen Farbdifferenz wird gesucht. Die Anzahl an Regionen zwischen den beiden plus eins ergibt die angenommene Anzahl der Farben im Muster entlang der Geraden. Es werden aus der Geraden Hintergrundregionen und Bereiche entfernt, die an den Rändern der Geraden liegen und von der betrachteten Region durch Hintergrundregionen oder einer langen Folge von Pixeln undefinierter Farbbereiche getrennt werden.

Aus allen betrachteten Geraden wird zur Bestimmung der Farbgruppen diejenige ausgewählt, deren Farbanzahl u minimal und größer eins ist (Abb. 6.4 c). Ausgehend vom betrachteten Farbbereich werden jeweils die Farbdifferenzen zu den $u - 1$ Regionen auf der einen und anderen Seite der Gerade bestimmt und jeweils addiert. Die Seite mit der größten Summe bestimmt die Farbgruppen, die als Zwischenregionen des Musters entlang der Gerade definiert werden. Diese werden in $\mathcal{R}$ aufgenommen.

Entlang der Gerade wird das Muster erweitert (Abb. 6.4 d und Abb. 6.5). Erstellt werden u Graphen, die als Knoten die Farbregionen enthalten, die entlang der Geraden u Positionen entfernt voneinander liegen (siehe Abb. 6.5). Pro Graph werden die Knoten mit Kanten verbunden, falls diese Regionen u Positionen voneinander entfernt liegen und einen hinreichend ähnlichen Farbwert aufweisen. Eine Kante zwischen zwei Knoten repräsentiert die dazwischen liegenden Farbbereiche anderer Farbgruppen. Sind diese Farbbereiche in deren Graphen von den Nachbarknoten getrennt worden, so wird deren zugehörige Kante von den anderen Graphen getrennt. Zu den Farbgruppen (siehe Abb. 6.5 gestrichelt) werden diejenigen hinzugefügt, die mit den Regionen in den Graphen über einen Weg erreicht werden können.

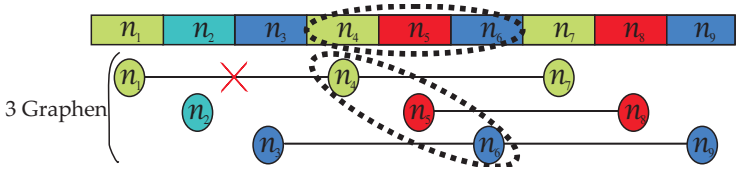


Abbildung 6.5: Beispiel zur Bestimmung des Musters entlang einer Gerade.

Das bestimmte Muster entlang der Gerade wird schließlich über das ganze Bild erweitert, falls eine Farbklasse mindestens zwei Regionen enthält und alle anderen mindestens eine (Abb. 6.4 e). Solange neue Farbregionen zu den Farbgruppen zugeordnet werden, wird der Graph $\mathcal{G}$ traversiert. Abwechselnd für jede Farbgruppe werden dessen direkte Nachbarn bestimmt. Nachbarn mit einer mittleren Farbe ähnlich zu der nächsten Farbgruppe werden in die nächste Farbgruppe aufgenommen. Hintergrundbereiche werden nur aufgenommen, falls diese von der Größe zum Muster passen. Gehört der Hintergrund zu einem Farbcluster, so werden die Voraussetzungen für die Farbcluster untersucht. Ist eine Farbregion n_k Teil eines Musters und eines Farbclusters Ξ_v, werden alle Regionen aus dem Cluster dem Muster hinzugefügt.

3.4 Verarbeitung der entstandenen Musterbereiche

In einem ersten Schritt werden alle Muster verworfen, die Teil eines größeren sind. Des Weiteren werden ähnliche Muster mit gleicher Anzahl an Farbgruppen, gemeinsamen Farbregionen und ähnlichen Farbgruppen verschmolzen.

In einem weiteren Schritt werden Farbregionen, die auf mehreren Mustern auftreten, einem einzelnen Muster zugeordnet. Priorität haben die Muster, die wegen einer höheren Frequenz früher aufgetreten sind. Aufgrund der eindeutigen Zuordnung ist es möglich, dass Farbregionen im Muster enthalten sind, die keine direkten Nachbarn im Muster haben. Diese Regionen werden aus dem Muster entfernt. Ebenfalls können Muster nicht mehr zusammenhängend sein. Diese werden auf zusammenhängende Muster verteilt. Als Frequenz der neuen Muster wird die maximale Frequenz der enthaltenen Farbregionen zugeordnet. Die Muster werden dann anhand ihrer Frequenzen sortiert. Schließlich

werden aus den Mustern diejenigen Farbbereiche entfernt, die nur ein „Rahmen" des Musters sind.

Da Kanten zwischen Farbregionen eine große Farbdifferenz aufweisen, werden diese nicht einer Farbregion zugeordnet. In einem letzten Schritt werden deswegen Pixel, deren Farben zwischen den Farbklassen der Musterregionen sind und direkt neben Musterregionen liegen, als „Übergangspixel" eingestuft und zu dem Muster hinzugefügt.

4 Ergebnisse

In Abbildung 6.6 sind detektierte Musterregionen sowie farbhomogene Hintergrundregionen des vorgestellten Detektors zu sehen. Die ersten zwei Bilder sind mit einem Raytracing-Programm erstellt worden und letzteres wurde selbst aufgenommen. Es ist zu erkennen, dass unterschiedliche Musterregionen unabhängig von ihrer Form, Farbe und Position in Bild erkannt werden. Weiterhin wird das Schachbrettmuster in dem mittleren Bild trotz Verdeckung durch die Kugeln, die ein anderes Muster besitzen, erkannt.

Die Anwendung des Detektors auf verschiedene Testbilder hat gezeigt, dass Musterregionen robust erkannt werden, wenn diese nicht ineinander verschachtelt sind und die unterschiedlichen Farbbereiche durch eindeutige, nicht verschwommene Grenzen getrennt werden.

5 Zusammenfassung

In diesem Beitrag wurde ein neuartiger verdeckungs- und affin-invarianter Regionendetektor vorgestellt. Dieser Detektor basiert auf Farb- und Frequenzinformation und ist in zwei Teile aufgegliedert. Der erste Teil liefert überschneidungsfreie, zusammenhängende Farbbereiche und im zweiten Teil werden die Musterkombinationen bestimmt. Die Ergebnisse zeigen, dass zusammengehörende Bildbereiche bei einer hinreichenden Bildqualität zuverlässig erkannt werden.

Literatur

1. J. Matas, O. Chum, M. Urban und T. Pajdla, „Robust wide-baseline stereo from maximally stable extremal regions", *Image and Vision Computing*, Vol. 22,

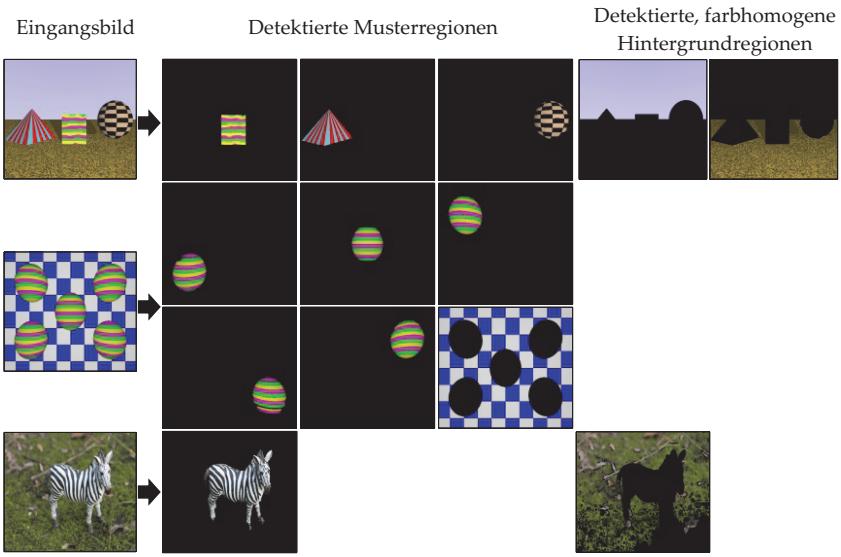


Abbildung 6.6: Ergebnisse des vorgestellten Detektors.

Nr. 10, S. 761–767, 2004.

2. H. Bay, A. Ess, T. Tuytelaars und L. Van Gool, „Speeded-up robust features (surf)“, *Computer vision and image understanding*, Vol. 110, Nr. 3, S. 346–359, 2008.
3. K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir und L. Van Gool, „A comparison of affine region detectors“, *International journal of computer vision*, Vol. 65, Nr. 1-2, S. 43–72, 2005.
4. P.-E. Forssén, „Maximally stable colour regions for recognition and matching“, in *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07.*, S. 1–8.
5. U. Kiencke, M. Schwarz und T. Weickert, *Signalverarbeitung : Zeit-Frequenz-Analyse und Schätzverfahren*. München: Oldenbourg, 2008.
6. J. Beyerer, F. Puente León und C. Frese, *Automatische Sichtprüfung : Grundlagen, Methoden und Praxis der Bildgewinnung und Bildauswertung*. Berlin: Springer Berlin, 2012.

Fast image super-resolution utilizing convolutional neural networks

Hubert Soyer[1] and Christian Osendorfer[1]

Technische Universität München, Fakultät für Informatik, Lehrstuhl für Robotik und Echtzeitsysteme, Boltzmannstraße 3, 85748 München

Abstract We introduce an architecture for image super-resolution that is based on a combination of Convolutional Neural Networks and Sparse Coding. Our method achieves a high computational efficiency by approximating sparse codes through a fast direct inference technique. We achieve state-of-the art results on images with large spatial extent. Our experiments illustrate that performing standard upsampling methods on latent representations is highly beneficial over applying them directly to the original spatial domain. Our results indicate that the proposed architecture can serve as a basis for future improvements in single-image super-resolution.

1 Introduction

Super-resolution in computer vision is generally defined as the process of increasing the spatial resolution of one or a series of images. We attempt single-image super-resolution by leveraging a database of low-high resolution image pairs to learn the parameters of the proposed sparse coding based model. In related recent work [1, 2], the dictionary based sparse coding method [3–5] was applied with great success to single-image super-resolution.

In previous work, each sparse code links two different kinds of dictionaries: One dictionary with low-resolution atoms (filters) and another dictionary containing the corresponding high-resolution versions of these atoms. These two dictionaries are then used to super-resolve low-resolution images to their high resolution counterparts in the following, straight forward way: First, the low-resolution image is split up into patches featuring the same spatial dimensions as the low-resolution

atoms. For each of these patches, its sparse code relative to the low-resolution dictionary is determined, i.e. the low-resolution patch can be expressed as a weighted sum over the low-resolution atoms where the weight for each of these atoms is the corresponding entry in the sparse code. A high-resolution version of the patch is then computed as a weighted sum of the high-resolution atoms, utilizing the sparse code learned on the low-resolution patch. Splitting the overall images into patches and then re-aligning the super-resolved patches introduces problems – Yang et al. [2] extend the idea by a global reconstruction constraint to counteract these problems. They further propose a version of their model specifically tailored towards face and natural images. Instead of upsampling images directly like e.g. bicubic interpolation based super-resolution [6], the previously mentioned approaches tackle this problem in an indirect way. They decompose the low-resolution image into atoms and recompose it from higher-resolution atoms. Because of the fixed sized atoms, images that are larger than a single patch require patch-wise processing. Patch-wise processing [7] typically causes issues at the patch borders and involves a lot of computational overhead. Determining an appropriate sparse code for each patch is a costly procedure and thus applying the above mentioned work to large images is difficult in practice. According to Yang et al. their model [2] takes approximately 30 seconds to enlarge a 85×86 image to 255×258 with reasonably chosen hyperparameters on a Core duo@1.83 Ghz with 2GB Ram. [8] gives an overview on recent super-resolution techniques and lists execution times for a 128×128 image for different upsampling factors. To make high-quality image super-resolution faster and therefore available for practical applications we propose an architecture that leverages recent insights into fast approximate sparse coding and exploits a natural characteristic of the convolutional operator. Our approach can be *trained on exemplary image patches and scaled up to arbitrarily sized test images without any additional cost*. We present our method and the necessary preliminary work in section 2. Experimental details and results are described in section 3. Section 4 concludes with a brief outlook on future work.

Related Work. Sparse coding based algorithms have been employed very successfully to induce good feature representations of natural im-

ages [3–5, 9]. Unfortunately, they are often hard to apply in practice because of issues regarding their computational complexity. In order to find the latent feature representation of an unseen image, sparse coding based models generally require (iteratively) solving an optimization problem. [10, 11] propose to approximate the solution of this optimization problem through a feed forward neural network, reducing the iterative process to only a single step. We base our work on a convolutional extension to this idea introduced in [12] that is capable of learning a richer set of more diverse features. Using an approximator neural network allows us to achieve an application speed that makes our architecture applicable in practice.

2 Approach

Since their introduction, Convolutional Neural Networks (CNNs) [13] have continuously been yielding new state-of-the-art results on a variety of computer vision and image processing tasks [14, 15]. An often overlooked property of the convolutional operator is that it can be applied to inputs of arbitrary size without having to re-learn its kernels. The size of data the CNN is applied to after training does therefore not depend on the sizes of the patches or images it was trained on.

This property is the basis of our approach, it allows learning the parameters of the proposed model on *patches* and applying them to *full images*. Exploiting this property, we lose the ability to apply the previously described, standard sparse coding approach for super-resolution. Instead of dictionary elements that implicitly cause an upscaling (low and high resolution dictionary) the parameters of our model are convolutional filters/kernels, hence, the straight forward approach of implicit upsampling is not possible anymore.

Upsampling in the image domain is a widely used approach that reduces the problem of super-resolution to learning an optimal deconvolutional operator [16]. Instead of performing the upsampling in the image domain directly, we propose to magnify the latent, sparse code representations instead. We hypothesize that upsampling the adaptively learned latent representation of an image leads to higher quality results than upsampling the image in its original spatial domain.

As mentioned earlier, we can train the convolutional part of our

model on small patches and later apply it to arbitrarily sized images due to a convenient quality of the convolutional operator. To preserve this property, the upsampling step itself has to be a *spatially local operation*, i.e. every upsampled pixel can only depend on a neighborhood around the pixel that it corresponds to in the low-resolution image.

2.1 Model

The general architecture of the proposed model is introduced in [17]. We will describe the most important aspects of the proposed method and give a brief overview over different choices for the upsampling method that is applied to the sparse codes. Mathematically, at the center of our approach is the following objective function, which is explained in detail subsequently:

$$\mathcal{L}(x^{(lr)}, x^{(hr)}, z, \mathcal{D}^{(d)}, \mathcal{D}^{(e)}, W) = \tag{7.1}$$
$$\underbrace{\sum_{k=1}^{K} \|z_k{}^\star - f(\mathcal{D}_k^{(e)} * x^{(lr)})\|_2^2}_{\text{encoder}} +$$
$$\underbrace{\frac{1}{2} \|x^{(hr)} - \sum_{k=1}^{K} \mathcal{D}_k^{(d)} * \widehat{W z_k}\|_2^2}_{\text{decoder}} + \underbrace{\lambda |z|_1}_{\text{sparsity}}$$

General Architecture. The proposed model contains 6 different sets of variables: (i) Input pixels (low-resolution, $x^{(lr)}$). (ii) Output pixels (high-resolution, $x^{(hr)}$). (iii) Sparse codes (z for all sparse codes or z_k for sparse code k). (iv) Decoder (sparse code $\rightarrow$ image) convolutional filters ($\mathcal{D}^{(d)}$, kernels) that are applied to the sparse codes. (v) A matrix W that performs the upscaling step. The matrix is multiplied with the flattened low-resolution sparse codes which results in flattened high-resolution sparse codes (i.e. $W \in \mathbb{R}^{o_{(hr)} \cdot p_{(hr)} \times o_{(lr)} \cdot p_{(lr)}}$). We can either regard the matrix as another set of parameters and optimize it regarding the training objective or set its values fixed to resemble a standard upsampling method. (vi) Encoder (image $\rightarrow$ sparse code) convolutional filters ($\mathcal{D}^{(e)}$) corresponding to the neural network that generates an approximation

of the sparse codes given an image. We formulate an objective function that allows us to jointly train the sparse coding architecture with the approximator (encoder). Equation 7.1 gives this objective function $\mathcal{L}$. The equation comprises three major parts. The decoder, the encoder and the sparsity term. Each of these parts constitutes an objective function on its own. We minimize the sum over all three sub-objectives in an iterative manner employing stochastic gradient descent.

The encoder sub-objective is computed as the norm over the distance between the guiding sparse codes (z^*) and their approximated version generated by the model. We produce the guiding sparse codes at each step of the iterative optimization process by applying the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [18], which is itself iterative. The iterative and therefore costly nature of FISTA is the reason why we approximate its results instead of using the algorithm directly. During training we utilize FISTA as a guide for the encoder part of the proposed model but replace it entirely by the encoder at application time.

Similarly to the encoder part, the decoder term also represents the norm over a difference. In this formulation we propose the l^2 norm to measure the distance between the target high-resolution image ($x^{(hr)}$) and the output of our model ($\sum_{k=1}^{K} \mathcal{D}_k^{(d)} * \widehat{W z_k}$). We will later show that other norms perform empirically better with respect to several standard error measures for super-resolution.

The third sub-objective represents the l^1-norm over all codes z of the currently processed image. Since the l^0 pseudo norm – which directly measures the sparsity of a vector or matrix – is computationally intractable, it is in practice often replaced by other l^q norms with $q \leq 1$. We chose the l^1 norm as it can be conveniently optimized through the combination of FISTA and the soft-thresholding function. Figure 7.1 illustrates the proposed architecture graphically, showing examples of intermediate values and parameters at several stages of the pipeline.

Sparse Code Upsampling. As mentioned previously, the sparse code upsampling Matrix W can either be learned as part of the optimization process or set to a fixed value. We compare the learned version with several fixed versions of the matrix. When learning the matrix, we don't optimize for all entries of the matrix. Instead we mask the matrix

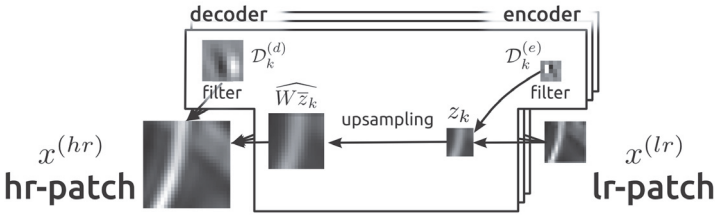


Figure 7.1: Graphical illustration of the model. The low-resolution input image is convolved with each of the encoder filters to create low-resolution codes. This code is then upsampled and convolved with the corresponding decoder filters. This yields $\#filters$ many partial reconstructions which are summed up with equal weight to acquire the high-resolution output image. The encoder-upscaling-decoder plate in the model illustrates the application of encoder convolution, sparse code upsampling and decoder convolution for one filter pair.

such that only the weights for the 4 neighbors of each pixel are learned. By introducing these locality constraints we ensure that the method is scalable i.e. we can apply the method to arbitrarily shaped images after learning its parameters on fixed sized patches.
In addition to standard bilinear and nearest neighbor upsampling we introduce *perforated* and *linear shifted* upsampling. The former only copies the pixels from the low-resolution image to their positions in the high-resolution image and leaves all other values blank. The latter is similar to bilinear interpolation with more weight on the upper left pixel and can therefore be regarded as a mix between nearest neighbor upsampling and bilinear interpolation based upsampling.

3 Dataset and Evaluation

A major merit of the proposed model is its speed. In contrast to most other recent sparse coding based super-resolution approaches [2,19] the proposed architecture can deal with images that feature very large spatial dimensions. To demonstrate the practical applicability of our approach we compare our model to bicubic spline based interpolation, bilinear interpolation and nearest neighbor interpolation on the Van Hateren dataset [20]. Due to the large size of the images in this dataset

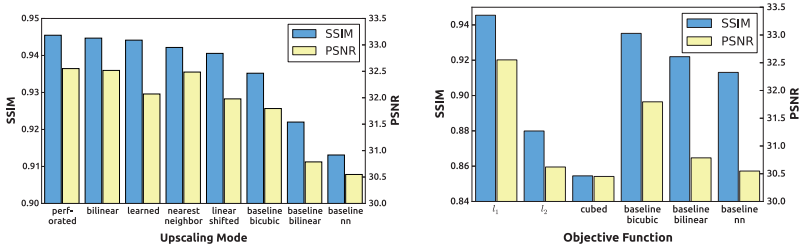
we chose only high-performance standard super-resolution algorithms as our baseline – it was computationally not feasible to apply the previously cited sparse coding based super-resolution algorithms.

The Van Hateren dataset contains 4167 gray scale nature and architecture pictures with 1536×1024 pixels each and a gray scale depth of 12 bit. Our training set constitutes 8000 randomly sampled 50×50 high resolution patches and their 25×25 down sampled low-resolution counterparts (factor 2). The patches were extracted at random positions from 400 images, 20 from each image. For building a validation set we sampled 2000 patch pairs from another 100 images in the same way. The test set consists of 100 unseen, *full-sized* images.

Following related literature [19, 21] we report results using the most commonly used error measures for super-resolution: The Peak-Signal-to-Noise Ratio (PSNR) and the Structured Similarity (SSIM), where the latter has been argued to resemble the human impression of reconstruction errors more closely. [17] only gives a very brief and general overview over the performance of our method. In this work, we will focus on several aspects that are crucial for the proposed architecture to work and provide insights into why certain design choices have a large impact on the performance of the model. We empirically investigate the following aspects:

- Comparison of the output quality between the learned upscaling matrix and several fixed settings.
- Effect of different norms in the decoder sub-objective.
- Relationship between different fixed upscaling methods and the influence of sparsity.
- Effect of different decoder and encoder filter size combinations.

Figure 7.2(a) shows a comparison of the standard high performance super resolution baselines and the proposed method with different choices for the upsampling method. Interestingly, *perforated*, the simplest type of upsampling, performs best. This method does not interpolate at all, it simply sets the low-resolution values at their corresponding positions in the high-resolution sparse code and leaves all remaining pixels in the high-resolution code blank (at zero). Applying *perforated* upsampling in practice is very efficient, it only requires iterating over all values of the low-resolution code once. Even though the *perforated* upscaling method



(a) Comparison of PSNR and SSIM for the baselines and several variants of the upscaling matrix W.

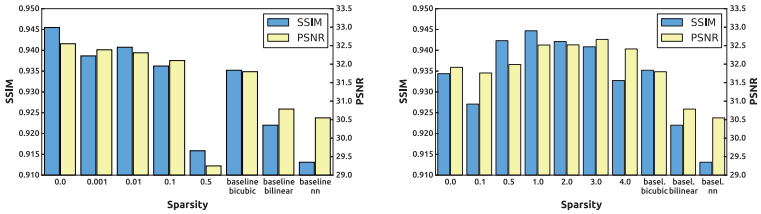
(b) Comparison of different norms as penalty in the decoder sub-objective. *cubed* refers to the l^3 norm.

Figure 7.2: Impact of decoder sub-objective norms and upscaling methods on performance (higher values are better).

leaves wide blank gaps in the upsampled sparse code, the final reconstruction performs best over all variants of W – all the blanks are filled without residue by the convolutional decoder. *Perforated* is followed by *bilinear* interpolation based upsampling. This interpolation based process is far more computationally expensive than *perforated* upsampling. It not only requires iterating over all pixels in the high instead of low-resolution image but also involves arithmetic operations instead of just plain copy operations.

The learned version of the upsampling matrix performs almost on-par with *perforated* and *bilinear* regarding SSIM but loses ground regarding the PSNR measure. We hypothesize that because of the stochastic nature of the optimization process it is difficult to learn the upsampling matrix jointly with the decoder and encoder filters and recommend to fall back to the simpler and computationally more efficient *perforated* upsampling for practical applications. On a modern Desktop CPU an unoptimized implementation of the proposed method took about 4 seconds to scale a full image from the Van Hateren dataset up by a factor of 2. Bicubic interpolation based upsampling took approximately 2.5 seconds.

The nearest neighbor and linear shifted approach (which could be viewed as a mix between linear interpolation and nearest neighbor) perform worst among all upscaling variations of the proposed model but still manage to beat the best baseline results by a margin.



(a) Influence of sparsity (λ) on perforated upscaling

(b) Influence of sparsity (λ) on bilinear upscaling

Figure 7.3: Relationship between perforated/bilinear upsampling and the influence of sparsity (higher values are better).

We highlight that the linear interpolation and the nearest neighbor upscaling *applied to the sparse codes* perform significantly better than the corresponding baselines. This supports our hypothesis that applying transformations in the latent sparse coding domain works considerably better than applying them in the original image domain.

We considered several different norms for the decoder objective functions to penalize reconstruction errors. As evident in figure 7.2(b) only with the l^1 norm in the decoder sub-objective the architecture manages to beat the baselines. The l^1 norm penalizes small differences ($\ll 1$) stronger than the l^2 norm and much stronger than the l^3 norm. Taking these small differences into account, however, seems to be crucial for a sharp reconstruction of the high-resolution image. We found that images processed using the l^2 or l^3 norm looked significantly blurrier resulting in a worse score.

Figure 7.3 shows the influence of sparsity using two different upsampling techniques. When the sparse codes are upsampled through *bilinear* interpolation (Figure 7.3(b)) the quality of the achieved results peaks when the sparsity sub-objective is roughly weighed equal to the other sub-objectives of the proposed overall objective function ($\lambda \approx 1.0$). Upsampling with the *perforated* setting exhibits a very different pattern, peaking at 0.0 sparsity influence and decreasing result quality as the influence of the sparsity sub-objective increases.

(a) *nearest neighbor*, *linear shifted* and *bilinear* upscaling

	nearest neighbor			**linear shifted**			**bilinear**		
encoder	5	5	5	5	5	5	5	5	5
decoder	9	10	11	9	10	11	9	10	11
PSNR	**32.49**	31.14	28.72	**32.17**	30.88	28.65	32.18	32.33	**32.52**
SSIM	**0.9422**	0.9222	0.8764	**0.9441**	0.9252	0.8794	0.9445	0.9385	**0.9447**

(b) *perforated* upscaling

	perforated								
encoder	5	5	5	7	7	7	9	9	9
decoder	9	10	11	13	14	15	17	18	19
PSNR	32.41	**32.55**	32.43	**32.34**	32.26	32.31	**32.32**	32.30	31.89
SSIM	0.9393	**0.9455**	0.9435	0.9404	0.9415	**0.9419**	**0.9433**	0.9413	0.9430

Table 7.1: PSNR and SSIM for several combinations of encoder and decoder filter sizes with different upscaling methods (higher values are better).

We hypothesize that this behavior is due to the sparse nature of the *perforated* upsampling – for every value in the low-resolution image, it creates three blank pixels in the high-resolution version, resulting in a very sparse upsampling output (three blank pixels for one non-blank pixel, assuming upscaling by a factor of two). While the bilinear upscaling mode does not inherently introduce sparsity, it benefits from the sparsity sub-objective instead. In both presented cases sparsity is crucial to achieve high result quality, it can be either a characteristic of the upscaling method or a sub-objective. Without having an upsampling layer between the encoder and decoder the sizes of both sets of filters would have to be equal. In the proposed architecture, however, these filter sizes are decoupled, enabling us to choose arbitrary sizes for either and making the choice of both sizes subject to hyperparameter tuning.

We empirically found that the best choice for the width of the quadratic decoder filters is close to [upscaling factor] $\cdot$ [width of encoder filter]. Deviating far from this rule decreased the output quality heavily. However, even for sizes within 1 pixel of the suggested value we found fluctuations. Table 7.1 illustrates that depending on the chosen upscaling these differences can be large. The *perforated* method (table 7.0(b)) is relatively robust to sizes of ± 1 pixels around the suggested value. On the contrary, for the methods shown in table 7.0(a), one pixel difference has a huge impact on SSIM and PSNR.

4 Conclusion and Future Work

We introduced a single image super-resolution architecture based on approximate convolutional sparse coding. Our approach achieves state-of-the-art results among methods suitable for large image super-resolution. We showed that upsampling in the sparse feature domain is superior compared to applying upsampling directly in the original image domain.

Promising results with linear techniques for sparse code upsampling clearly motivate the exploration of more complex methods like bicubic spline based upsampling. Furthermore, we identified the encoder convolutional neural network (approximating FISTA) as another potential quality bottleneck that should be considered in future work. The work in [22], proposing an architecture based on unrolling FISTA, poses an interesting starting point for further improvements in this direction.

References

1. F. Couzinie-Devy, J. Mairal, F. Bach, and J. Ponce, "Dictionary learning for deblurring and digital zoom," *CoRR*, 2011.
2. J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *Image Processing, IEEE Transactions on*, vol. 19, no. 11, 2010.
3. B. J. Olshausen and D. J. Field, "Sparse coding with an over complete basis set: a strategy employed by v1?" *Vision Research*, 1997.
4. J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *CVPR*, 2008.
5. H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *NIPS*, 2007.
6. M. Petrou and C. Petrou, *Image Processing: The Fundamentals*. Wiley & Sons, 2010.
7. Z. Cui, H. Chang, S. Shan, B. Zhong, and X. Chen, "Deep network cascade for image super-resolution," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 49–64.
8. C.-Y. Yang, C. Ma, and M.-H. Yang, "Single-image super-resolution: A benchmark," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 372–386.

9. K. Yu, Y. Lin, and J. Lafferty, "Learning image representations from the pixel level via hierarchical sparse coding," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1713–1720.
10. K. Kavukcuoglu, M. Ranzato, and Y. LeCun, "Fast inference in sparse coding algorithms with applications to object recognition," Computational and Biological Learning Lab, Courant Institute, NYU, Tech. Rep., 2008.
11. K. Kavukcuoglu, M. Ranzato, R. Fergus, and Y. Le-Cun, "Learning invariant features through topographic filter maps," in *CVPR*, 2009.
12. K. Kavukcuoglu, P. Sermanet, Y.-L. Boureau, K. Gregor, M. Mathieu, and Y. L. Cun, "Learning convolutional feature hierarchies for visual recognition," in *NIPS*, 2010.
13. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, vol. 86, 1998.
14. D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 3642–3649.
15. P. Pinheiro and R. Collobert, "Recurrent convolutional neural networks for scene labeling," in *Proceedings of The 31st International Conference on Machine Learning*, 2014, pp. 82–90.
16. C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 184–199.
17. C. Osendorfer, H. Soyer, and P. van der Smagt, "Image super-resolution with fast approximate convolutional sparse coding," in *ICONIP*, 2014.
18. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
19. L. He, H. Qi, and R. Zaretzki, "Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution," in *CVPR*, 2013.
20. J. H. v. Hateren and A. v. d. Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proceedings: Biological Sciences*, vol. 265, no. 1394, 1998.
21. K. Zhang, X. Gao, D. Tao, and X. Li, "Multi-scale dictionary for single image super-resolution," in *CVPR*, 2012.
22. D. Sprechmann, A. M. Bronstein, and G. Sapiro, "Learning efficient sparse and low rank models." *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 2012.

Optimierung der *Fast Radial Symmetry Detection* für eine echtzeitfähige Kreisdetektion

Stefan Eickeler[1] und Matias Valdenegro[2]

[1] Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS,
stefan.eickeler@iais.fraunhofer.de
Schloss Birlinghoven, 53754 Sankt Augustin
[2] Hochschule Bonn-Rhein-Sieg
Grantham-Allee 20, 53757 Sankt Augustin

Zusammenfassung Die Hough-Transformation kann zur Detektion von Linien, Kreisen und beliebigen Formen eingesetzt werden. Unter dem Namen *Fast Radial Symmetry Detection* wurde von Loy und Zellinsky eine Erweiterung der Hough-Transformation vorgestellt, die radialsymmetrische Objekte detektieren kann. Hierdurch ist sie besonders gut zur schnellen Detektion von Kreisen mit Radien innerhalb eines vorgegebenen Intervalls geeignet. In dieser Veröffentlichung wird dieses Verfahren in die Gruppe deren Hough-Transformation-basierten Verfahren eingeordnet und auf Möglichkeiten zur Geschwindigkeitsoptimierung untersucht. Die Verwendbarkeit des Verfahrens zur Detektion von Verkehrszeichen wird anhand des Datensatzes des *German Traffic Sign Detection Benchmark* verifiziert.

1 Einleitung

Die Detektion von Kreisen in Bildern ist eine weit verbreitete Problemstellung in der Bildverarbeitung. Für die Lösung dieses Problems gibt es bereits verschiedene Algorithmen [1–5]. Neben der Anforderung nach einer hohen Detektionsrate ist die schnelle Berechenbarkeit eine der wichtigsten Anforderungen an den Kreisdetektor.

Zunächst wird die Entwicklung innerhalb der Hough-basierten Verfahren zur Kreisdetektion betrachtet. Anschließend wird das Originalverfahren beschrieben, da es in mehreren aufeinander aufbauen-

den Veröffentlichungen beschrieben ist und Raum für Interpretationen bezüglich der exakten Implementierung lässt.

2 Hintergrund

Die Hough-Transformation zur Detektion von Linien in Bildern geht auf das Patent [6] von Paul Hough aus dem Jahr 1962 zurück. Die Entstehungsgeschichte der Hough-Transformation für Linien ist in [7] ausführlich dargestellt.

Die erste Erweiterung der Hough-Transformation zur Kreisdetektion wurde in [1] vorgestellt. Hier wurde das zweidimensionale Akkumulatorarray (Steigung und Offset bzw. Winkel und Abstand) der Liniendetektion auf ein dreidimensionales Akkumulatorarray (Mittelpunkt und Radius) erweitert. Für Pixel im Bild, die Teil eines Kreises sein können, werden alle Akkumulatorzellen inkrementiert, welche Kreise repräsentieren, die diesen Pixel schneiden. Anschließend werden die Akkumulatoarrays geglättet und eine Maximumsdetektion durchgeführt.

In [2] wird von Kimme, Ballard und Sklansky die Rechenzeit der Detektion durch die Nutzung der Gradienten der Kanten innerhalb des Bildes reduziert. Es werden nur die Akkumulatorzellen inkrementiert, welche Kreise repräsentieren deren Mittelpunkt in Gradientenrichtung liegt.

In [8] wird von Ballard eine Erweiterung der Hough-Transformation auf beliebige Formen vorgestellt. Es ist auch die erste Veröffentlichung, die die Äquivalenz von einer Gaußfilterung nach der Transformation und einer Gaußfilterung während der Inkrementierung der Akkumulatorzellen aufzeigt.

In [9] wird das Verfahren von [2] erweitert, so dass nicht nur Zellen im Akkumulatorarray inkrementiert werden, sondern auch die Zellen in negative Gradientenrichtung dekrementiert werden. Hierdurch hebt sich der Einfluss vom Bildrauschen im Akkumulatorarray gegenseitig auf und die Bestimmung des Schwellwertes für die Kreisdetektion wird deutlich vereinfacht.

In [10] vereinfachen Barnes, Zelinsky und Fletcher das Verfahren explizit für die Detektion von Kreisen.

3 *Fast Radial Symmetry Detection*

Der *Radial Symmetry Detector* (RSD) ist ein Gradienten-basierter Hough-ähnlicher Algorithmus, der Kreise mit bekannten Radien innerhalb der Menge $r \in R$ detektieren kann. Die folgende Beschreibung wurde aus den Veröffentlichungen [9,10] und dem von den Autoren bereitgestellten Matlab-Code zusammengestellt. Abbildung 8.1.1 stellt die Verarbeitungsschritte dar. Ausgehend von dem Eingangsbild wird als erstes ein (a) Gauß'sches Filter mit sehr kleinem Filterradius zur Weichzeichnung eingesetzt. Im folgenden Schritt (b) wird mit dem Sobel-Operator der Gradient berechnet und in die Kantenintensität (Betrag) und Ausrichtung (Einheitsvektor) umgewandelt. Die Mittelpunktschätzung (c) hat für jeden zu detektierenden Kreisradius jeweils ein zweidimensionales Akkumulatorarray O_r mit jeweils zum Eingangsbild identischer Auflösung. Jeder Bildpunkt, dessen Gradientenbetrag oberhalb einer Schwelle G_t liegt, wird als Teil eines Kreises angenommen. Für jeden zu detektierenden Kreisradius werden jeweils jene Akkumulatorzellen geändert, deren Positionen folgendermaßen definiert sind:

$$\mathbf{p}_+ = \mathbf{p} + \left\lfloor r\frac{\mathbf{g}(\mathbf{p})}{||\mathbf{g}(\mathbf{p})||} + \begin{pmatrix}0,5\\0,5\end{pmatrix}\right\rfloor \tag{8.1}$$

$$\mathbf{p}_- = \mathbf{p} - \left\lfloor r\frac{\mathbf{g}(\mathbf{p})}{||\mathbf{g}(\mathbf{p})||} + \begin{pmatrix}0,5\\0,5\end{pmatrix}\right\rfloor \tag{8.2}$$

Wobei $\mathbf{g}(\mathbf{p})$ der Gradient des Pixels an Position $\mathbf{p}$ ist. $\lfloor\ \rfloor$ ist die Abrundungsfunktion, die bei jedem Element des Vektors die Nachkommastellen abschneidet. Für jeden Gradientenpixel werden zwei Änderungen

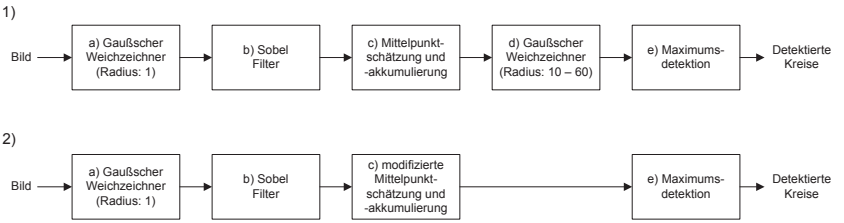


Abbildung 8.1: Originalverfahren (1) und optimiertes Verfahren (2) zur Kreisdetektion.

vorgenommen, um die Gradientenrichtungen dunkel nach hell und hell nach dunkel zu berücksichtigen:

$$O_r(\mathbf{p}_+) = O_r(\mathbf{p}_+) + 1 \tag{8.3}$$

$$O_r(\mathbf{p}_-) = O_r(\mathbf{p}_-) - 1 \tag{8.4}$$

In Schritt (d) wird jeweils das *symmetry contribution image* F_r aus dem Akkumulatorarray O_r berechnet und F_r mit einem Gauß'schen Filter A_r geglättet.

$$F_r(\mathbf{p}) = \left(\frac{O_r}{k_r}\right)^{\alpha} \tag{8.5}$$

$$S_r = F_r * A_r \tag{8.6}$$

Wobei k_r eine radiusabhängige Normalisierungskonstante und α der *radial strictness* Parameter ist, der den empfohlenen Wert $\alpha = 2$ besitzt [9]. Das Gauß'sche Filter hat die Größe $r \times r$ mit der Standardabweichung $\sigma = 0{,}5\,r$. Dieser Rechenschritt benötigt ungefähr 90 % der Gesamtrechenzeit und hat daher das größte Optimierungspotential. Als abschließender Schritt (e) wird eine Detektion von Betragsmaxima in den gefilterten Akkumulatorarrays S_r durchgeführt. Ein Maximum gibt den Mittelpunkt des Kreises bei dem vorgegebenen Radius an. Das Ergebnis sind die detektierten Kreise im Bild.

In Abbildung 8.2.1 sind die Akkumulatorarrays für ein Beispielbild dargestellt.

4 Optimierte Version der *Fast Radial Symmetry Detection*

Im Vergleich mit dem Standardverfahren kann das optimierte Verfahren ohne die rechenintensive Gaußfilterung (d) mit großen Filterradien arbeiten.

5 Beschreibung des Verfahrens

In Abbildung 8.1.2 ist die verbesserte Prozesskette dargestellt. Die Verarbeitungsschritte (a) und (b) sind identisch mit dem Original verfahren in Abbildung 8.1.1.

Abbildung 8.2: Visualisierung der Akkumulatorarrays für das Originalverfahren (1) und das optimierte Verfahren (2).

Anstelle von Akkumulatorarrays mit konstanter Größe haben die Arrays eine Größe die der Originalgröße multipliziert mit $\frac{r_{\text{base}}}{r}$ entspricht. r_{base} ist typischerweise der kleinste zu detektierende Radius. Hierdurch wird der Speicherbedarf der Akkumulatorarrays ungefähr halbiert. In der Mittelpunktschätzung (c) werden jeweils die 2×2 Akkumulatorzellen, die sich aus den korrigierten Pixelposition plus (bzw. minus) dem Einheitsvektor ergibt, um eins erhöht (bzw. erniedrigt):

$$\mathbf{p}_+ = \left\lfloor r_{\text{base}} \left(\frac{1}{r}\mathbf{p} + \frac{\mathbf{g}(\mathbf{p})}{||\mathbf{g}(\mathbf{p})||} \right) \right\rfloor \tag{8.7}$$

$$\mathbf{p}_- = \left\lfloor r_{\text{base}} \left(\frac{1}{r}\mathbf{p} - \frac{\mathbf{g}(\mathbf{p})}{||\mathbf{g}(\mathbf{p})||} \right) \right\rfloor \tag{8.8}$$

$$\begin{aligned}
O_r\left(\mathbf{p}_+\right) &= O_r\left(\mathbf{p}_+\right) + 1 \\
O_r\left(\mathbf{p}_+ + \begin{pmatrix}1\\0\end{pmatrix}\right) &= O_r\left(\mathbf{p}_+ + \begin{pmatrix}1\\0\end{pmatrix}\right) + 1 \\
O_r\left(\mathbf{p}_+ + \begin{pmatrix}0\\1\end{pmatrix}\right) &= O_r\left(\mathbf{p}_+ + \begin{pmatrix}0\\1\end{pmatrix}\right) + 1 \\
O_r\left(\mathbf{p}_+ + \begin{pmatrix}1\\1\end{pmatrix}\right) &= O_r\left(\mathbf{p}_+ + \begin{pmatrix}1\\1\end{pmatrix}\right) + 1
\end{aligned} \tag{8.9}$$

$$\begin{aligned}
O_r\left(\mathbf{p}_-\right) &= O_r\left(\mathbf{p}_-\right) - 1 \\
O_r\left(\mathbf{p}_- + \begin{pmatrix}1\\0\end{pmatrix}\right) &= O_r\left(\mathbf{p}_- + \begin{pmatrix}1\\0\end{pmatrix}\right) - 1 \\
O_r\left(\mathbf{p}_- + \begin{pmatrix}0\\1\end{pmatrix}\right) &= O_r\left(\mathbf{p}_- + \begin{pmatrix}0\\1\end{pmatrix}\right) - 1 \\
O_r\left(\mathbf{p}_- + \begin{pmatrix}1\\1\end{pmatrix}\right) &= O_r\left(\mathbf{p}_- + \begin{pmatrix}1\\1\end{pmatrix}\right) - 1
\end{aligned} \tag{8.10}$$

Durch die Veränderung von 2×2 Akkumulatorzellen in Kombination mit den skalierten Akkumulatorarrays kann gegenüber dem Originalverfahren mit nur der Änderung von einer (1×1) Akkumulatorzelle der

Verarbeitungsschritt (d) entfallen, da eine Äquivalenz zwischen einer Gaußfilterung nach der Transformation und einer Gaußfilterung (hier durch 2×2 Pixel angenähert) während der Inkrementierung der Akkumulatorzellen existiert [8]. Die Formel 8.5 wird mit den Werten $\alpha = 1$ und $k_r = r$ vereinfacht angewendet. Für die Detektion der Maxima (e) werden jeweils drei benachbarte Radiusstufen für einen potentiellen Kreismittelpunkt zusammengefasst und gegen einen Schwellwert getestet. Durch die Zusammenfassung können auch Kreisradien, die zwischen den Radienstufen liegen zuverlässig erkannt werden.

5.1 Parameter

Das Intervall der Radien ist durch den Größenbereich der zu detektierenden Kreise vorgegeben. Im vorliegenden Fall ist der minimale Radius 10 Pixel. Dieser Radius wurde daher auch als Basisradius r_{base} festgelegt. Unter Vernachlässigung von Rundungsfehlern ergibt dieses bei 2×2 Inkrementierung eine Toleranz von ± 1 Pixel $\equiv \pm 10\,\%$ für jede Radiusstufe. Der Skalierungsfaktor für die folgende Radienstufe errechnet sich aus:

$$f = \frac{1+t}{1-t} \tag{8.11}$$

Wobei t die Toleranz und f der Skalierungsfaktor ist. Für einen Startradius von 10 Pixeln ergibt sich daher folgende Reihe:

Radiusstufe	1	2	3	4	5	6	7	8
Radius in Pixel	10,00	12,22	14,94	18,26	22,32	27,27	33,33	40,74

Der Gradientenschwellwert G_t liegt bei unseren Experimenten bei 30. Barnes et al. [10] zeigen, dass mit höheren Werten von G_t eine höhere Verarbeitungsgeschwindigkeit erreicht werden kann. Es ist aber davon auszugehen, dass höhere Werte die Detektionsrate reduzieren. Der Schwellwert für die Detektion der Betragsmaxima muss experimentell bestimmt werden.

6 Evaluierung

In Tabelle 8.1 wird ein erster Geschwindigkeitsvergleich verschiedener Implementierungen der RSD aufgelistet. Der Originalalgorithmus er-

Paper	Implementation	Resolution	Frames / sec	Pixels / sec
Barnes, Zelinsky, Fletcher, 2008 [10]	Multi-processing MMX	320 x 240	30 fps	2,3 M
Glavtchev, Muyan-Özcelik, Ota, Owens, 2011 [11]	GPU	640 x 480	33 fps	10,1 M
this approach, 2014	C++ single threaded	800 x 600	54 fps	25,9 M

Tabelle 8.1: Vergleich der Verarbeitungsgeschwindigkeiten unterschiedlicher RSD Implementierungen.

reicht eine Verarbeitungsgeschwindigkeit von 2,3 M Pixel/s. Der optimierte Algorithmus erreicht ohne Parallelisierung eine Verarbeitungsleistung von 25,9 M Pixel/s. Hierbei wurde durch die Verwendung einer älteren Notebook CPU versucht die Hardwarevoraussetzungen anzugleichen. Dennoch ist ein exakter Vergleich anhand der Tabelle 8.1 nicht möglich.

Für einen genaueren Vergleich der Algorithmen, wurden beide Algorithmen in C++ implementiert. Der Code für beide Algorithmen wurde mit den gleichen Techniken manuell optimiert. Es wurde keine Parallelisierung oder Vektorisierung der Algorithmen durchgeführt .

Für die folgenden Vergleiche wurden Bilder mit jeweils einem Kreis generiert. Der Radius des Kreises liegt innerhalb der Intervalls 8 bis 56 und damit im Erkennungsbereich der Detektoren. Den Bildern wurde additiv ein Gauß'sches Rauschen mit angegebener Standardabweichung hinzugefügt.

In Abbildung 8.3 wird die Laufzeit der beiden Algorithmen in Abhängigkeit von dem überlagerten Rauschen dargestellt. Wenn das Rauschen unterhalb des Gradientenschwellwertes liegt, wird die Rechenzeit des Originalverfahrens fast ausschließlich durch die Zeit für die Gaußfilterung (Abbildung 8.1 1.d) bestimmt. Das verbesserte Verfahren ist hier deutlich schneller, da dieser Verarbeitungsschritt entfällt. Sobald das Bildrauschen oberhalb des Gradientenschwellwertes liegt, steigt die Verarbeitungszeit für beide Verfahren an. Die Verarbeitungszeit des verbesserten Verfahrens steigt etwas stärker an als beim Originalverfahren, da 2 x 2 Pixel anstelle von nur einem Pixel in den Akkumulatorarrays geändert werden.

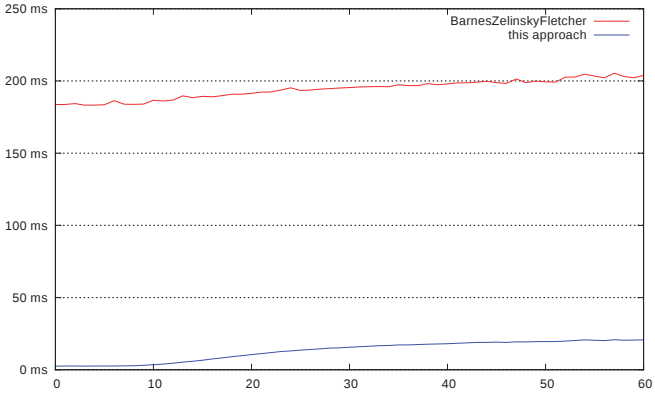


Abbildung 8.3: Rechenzeitbedarf des Originalverfahrens und des optimierten Verfahrens zur Kreisdetektion in Millisekunden abhängig vom überlagerten Rauschen in RMS.

Ein Vergleich der beiden Implementierungen auf realen Bildern zeigte, dass eine Geschwindigkeitsverbesserung um den Faktor 13 durch den verbesserten Algorithmus gegenüber der Originalversion bei gleicher Erkennungsleistung erreicht wird.

Beim Originalverfahren ist die Größe der Akkumulatorarrays für jeden Radius konstant. Hierdurch kann unabhängig vom Kreisdurchmesser eine pixelgenaue Lokalisierung des Kreises erfolgen. Beim verbesserten Verfahren ist die Größe der Akkumulatorarrays mit dem reziproken Wert der Radien skaliert. Durch diese Änderung ergibt sich eine Lokalisierungsgenauigkeit die relativ zum Kreisdurchmesser konstant ist.

In Abbildung 8.4 ist die Lokalisierungsgenauigkeit der beiden Verfahren abhängig vom Kreisradius dargestellt. Für beide Verfahren verhält sich die Genauigkeit wie erwartet. Beim verbesserten Verfahren sind zusätzlich Ungenauigkeitsspitzen an den Positionen der exakten Detektionsradien sichtbar. Diese konnten bei realen Anwendungen noch nicht beobachtet werden und müssen noch weiter untersucht werden.

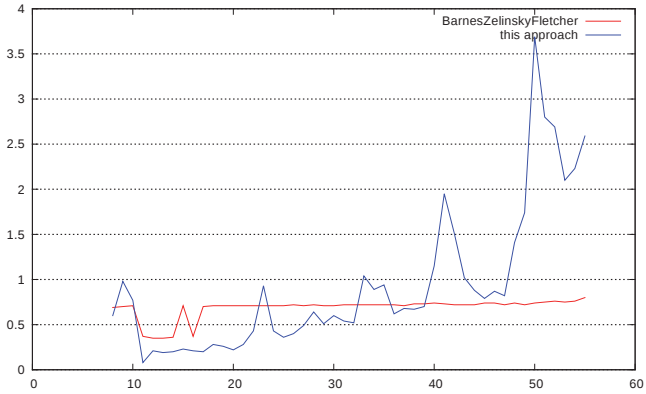


Abbildung 8.4: Lokalisierungsgenauigkeit des Originalverfahrens und des optimierten Verfahrens zur Kreisdetektion abhängig vom Kreisradius.

7 Verkehrszeichenerkennung mit der *Radial Symmetry Detection*

Die RSD kann mit einer nachgeschalteten Erkennungsstufe zu einer Verkehrszeichenerkennung für runde Verkehrszeichen ausgebaut werden (siehe Abbildung 8.5). Convolutional Neural Networks basieren auf dem Ansatz der rezeptiven Felder und haben sich als leistungsfähiges Verfahren zur Verkehrszeichenerkennung herausgestellt [12]. In diesem Fall wird eine einfache Netzwerkarchitektur (siehe Abbildung 8.6) verwendet, um das Verkehrszeichen zu erkennen und andere kreisförmige Objekte zurückzuweisen. Für das Training des Netzwerkes wurden annotierte Videoaufzeichnungen von Testfahrten verwendet.

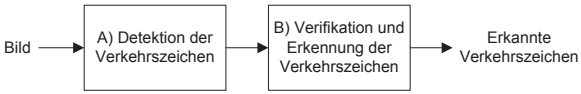


Abbildung 8.5: Verkehrszeichenerkennung basierend auf der optimierten RSD.

Die Evaluierung der Verkehrszeichenerkennung wurde auf dem Testdatensatz des *German Traffic Sign Detection Benchmark* (GTSRDB) [13]

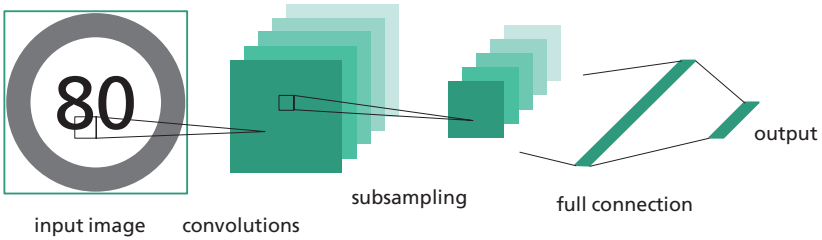


Abbildung 8.6: Schematische Darstellung des Convolutional Neuronalen Netzwerks für die Verifikation und Erkennung.

Precision	Recall	Erkennungsrate
99,4 %	96,3 %	97,5 %

Tabelle 8.2: Erkennungsergebnisse auf dem Testdatensatz des GTSRDB.

durchgeführt. Der Testdatensatz besteht aus 300 annotierten Bildern von Verkehrsszenen, die durch eine im Fahrzeug montierte Kamera aufgenommen wurden. Die Detektions- und Erkennungsleistung ist in Tabelle 8.2 dargestellt.

8 Zusammenfassung

Es wurde eine Laufzeitoptimierung der *Fast Radial Symmetry Detection* vorgestellt, die eine 13-fache Verareitungsgeschwindigkeit ermöglicht. Anhand von Experimenten wurde gezeigt, dass eine durch die Optimierung verursachte Lokalisierungsungenauigkeit für die praktische Anwendung keine Nachteile besitzten. Zukünftige Arbeiten werden sich mit der Erweiterung des Verfahrens auf beliebige Objektformen befassen.

Literatur

1. R. O. Duda und P. E. Hart, „Use of the Hough transformation to detect lines and curves in pictures“, *Commun. ACM*, Vol. 15, Nr. 1, S. 11–15, Jan. 1972.

2. C. Kimme, D. Ballard und J. Sklansky, „Finding circles by an array of accumulators“, *Commun. ACM*, Vol. 18, Nr. 2, S. 120–122, Feb. 1975.
3. M. Nixon, „Circle extraction via least squares and the Kalman filter“, in *Computer Analysis of Images and Patterns*, Ser. Lecture Notes in Computer Science, D. Chetverikov und W. Kropatsch, Hrsg. Springer Berlin Heidelberg, 1993, Vol. 719, S. 199–207.
4. M. Smereka und I. Duleba, „Circular object detection using a modified Hough transform“, *International Journal of Applied Mathematics and Computer Science*, Vol. 18, Nr. 1, S. 85–91, Mar. 2008.
5. L. Pan, W.-S. Chu, J. Saragih, F. De la Torre und M. Xie, „Fast and robust circular object detection with probabilistic pairwise voting“, *Signal Processing Letters, IEEE*, Vol. 18, Nr. 11, S. 639–642, Nov. 2011.
6. P. V. C. Hough, „Method and means for recognizing complex patterns“, Dec. 1962, US Patent 3,069,654.
7. P. Hart, „How the Hough transform was invented“, *Signal Processing Magazine, IEEE*, Vol. 26, Nr. 6, S. 18–22, Nov. 2009.
8. D. Ballard, „Generalizing the Hough transform to detect arbitrary shapes“, *Pattern Recognition*, Vol. 13, Nr. 2, S. 111 – 122, 1981.
9. G. Loy und A. Zelinsky, „Fast radial symmetry for detecting points of interest“, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 25, Nr. 8, S. 959 – 973, Aug. 2003.
10. N. Barnes, A. Zelinsky und L. Fletcher, „Real-time speed sign detection using the radial symmetry detector“, *Intelligent Transportation Systems, IEEE Transactions on*, Vol. 9, Nr. 2, S. 322–332, Jun. 2008.
11. V. Glavtchev, P. Muyan-Ozcelik, J. Ota und J. Owens, „Feature-based speed limit sign detection using a graphics processing unit“, in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, Jun. 2011, S. 195 –200.
12. P. Sermanet und Y. LeCun, „Traffic sign recognition with multi-scale convolutional networks“, in *Neural Networks (IJCNN), The 2011 International Joint Conference on*, Jul. 2011, S. 2809–2813.
13. S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing und C. Igel, „Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark“, in *International Joint Conference on Neural Networks*, Nr. 1288, 2013.

Exploitation of GPS control points in low-contrast IR imagery for homography estimation

Patrick Dunau

Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB),
Gutleuthausstr. 1, D-76275 Ettlingen

Abstract This paper gives a novel approach to homography estimation. It concentrates on the problem when there are no distinct features that can be matched between two consecutive images. The proposed method exploits knowledge of the camera system and sensors located on an aerial platform that is used to capture the infrared imagery. Exploiting additional information from a GPS receiver, an inertial measurement unit, and a compass, the position and direction of the camera can be determined. Once the information is available the projection mapping can be computed and real point targets can be inserted into the individual video frames. This technique yields specific point correspondences with which the homographies can be computed. The resulting homographies describe an improvement compared to homographies based solely on image features. The proposed method will be compared with sophisticated methods working with image features, e.g. gray values based methods. The aim of the method given in this paper is to overcome problems like drifting and the ambiguity of the pixel movements.

1 Introduction

This paper concentrates on a specific type of IR imagery which relies on scenes captured by a helicopter driven platform with an attached sensor platform containing a mid-wave IR (3-5 μm) camera and a long-wave IR (8-12 μm) camera. The sensors are used to capture objects and the surrounding area while approaching the scene. Mostly, the approach starts

from a long distance (≈24 km). Registering the images by computing the homographies between every image of the videos is inevitable if one used the videos for the purpose of signature evaluations, e.g. to evaluate the quality of a camouflage measure. Together with the small resolution of the cameras and due to strong noise effects image registration is a quite difficult task using these images.

The difficulty of finding point correspondences results in the complicated calculation of the homographies between two consecutive video frames. The first step is to obtain interest points from both images for which the homography is to be computed. Most registration methods rely on sophisticated methods like the Morevec operator [1], the Harris corner detection method [2], and the Foerstner operator [3]. The method proposed in this paper relies on the projection of control points into the images in order to have interest points. In the next step the determination of corresponding points in the two images is to be done. Classic methods like the correlation of gray values, or correlation in the fourier spectrum [4], or the wavelet transform [5] are used. Due to the noise effects the task is quite difficult. From the projection of the control points the point correspondences are naturally given. Having noisy point correspondences as input for algorithms computing a homography mapping results in erroneous mapping matrices. The computed homographies cause drifting effects when translating marked points throughout a video which results in quite big displacements when introducing markings early in the video. Also inconsistencies like sudden camera movements can happen. Mainly algorithms like RANSAC, or OpenCV's findHomographies are used to compute the homographies.

Thanks to the specific nature of the videos used here certain information is available. This includes the GPS position of the camera for each video frame as well as the camera parameters like sensor size and focal length. Together with this information the projection matrix for the camera can be computed. The destination area is well known. Known GPS-points are located in the destination area and in its surrounding area. Using the camera projection one can project specific GPS-points in general position into the camera images having good corresponding points in each image of the video. The projected GPS-points suffer only from the inaccuracy of the GPS-device and are not affected by the noise of the imaging system. So the drifting effect and inconsistencies can be overcome.

In section 2 the computation of the camera projection, as well as the projection of GPS control points is described. Also the estimation procedure of the homographies is presented. Results are shown in section 3 together with comparisons against other homography estimators. The last section discusses the presented procedure and gives hints for future research in this field.

2 Homography estimation using GPS control points

In order to perform homography estimation with the use of projected GPS control points, the available information has to be identified. The measurement system in use is called the AirSig-Platform. It consists of two IR cameras (MWIR/LWIR), a laser range finder, a daylight video camera, an inertial measurement unit, and a GPS device. For our purpose we use the images of the IR cameras, as well as the measurements of the GPS device. The parameters of the camera are well known. Given this information the camera projection matrix can be computed.

Formulation of the camera projection

The camera projection matrix is a mapping from the world coordinate system into the image coordinate system [6]. For the computation of the projection matrix the following parameters are required: The focal length of the camera system, as well as chip and pixel size. Also the resolution of the camera has to be known. To deduce the mapping from world coordinates into the image coordinate system one has to specify the position of the camera for each time step, which makes the mapping time variant. The time varying component only affects the position of the camera in world coordinates. The parameters are given in table 9.1.

Detector elements	640 x 512
Active area	20 μm x 20 μm
Focal length f	100 mm

Table 9.1: Camera parameters.

The position of the camera is given in GPS coordinates in the WGS84 coordinate system [7]. These coordinates are converted into the Earth

Centered Earth Fixed (ECEF) coordinate system, using the reference ellipsoid of the WGS84 coordinate system. Now, given the parameters from table 9.1 in combination with the camera coordinates in the ECEF coordinate system the projection matrix can be formed (9.1).

$$P_{cam} = K\,[I|-C] = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \left[\begin{array}{ccc|c} 1 & 0 & 0 & -x \\ 0 & 1 & 0 & -y \\ 0 & 0 & 1 & -z \end{array}\right] \tag{9.1}$$

Here f denotes the focal length of the camera, which is fixed, p_x and p_y are the respective coordinates of the principal point, i.e. the intersection of the optical axis with the image plane, in image coordinates. Together those parameters form the calibration matrix K. The negative x, y, and z coordinates of the camera position turns it into the world coordinate origin.

Projection of GPS control points

In order to project GPS control points into the images for homography estimation, the position of the camera has to be known for each frame of the video stream. The camera system is equipped with a GPS device which gives positional measurements at a specific rate (100 Hz). Consequently the true GPS positions of the camera have to be interpolated. A linear interpolation scheme can be used due to the fact that the helicopter moves constantly in the 40 ms between two frames. Once all GPS positions are computed, the camera matrices for each video frame can be constructed by replacing the $-C$ part of the matrix with the respective positions of the camera in converted ECEF-coordinates.

Given the camera matrices P_{cam} for each video frame, specific reference points have to be chosen in order to project them into each video frame. Figures 9.1 (a) and (b) give the same reference points in the first and last frame of the video.

Homography estimation

In order to estimate the homography, i.e. the projective mapping from image to image, one needs a sufficient number of point correspondences between the two images. From literature [6] it is known that at least 4

(a) (b)

Figure 9.1: (a) GPS reference point in the first frame of the video, (b) GPS reference point in the last frame of the video.

point correspondences are sufficient to fully determine the 8 degrees of freedom of the projective mapping.

Once the GPS reference points are placed in each frame of the video, the point correspondence problem can be considered as solved. This is due to the fact that the same GPS control points are projected into each video frame, which means there is a natural connection between the projected control points. Having thus acquired the point correspondences the homography estimation can now be performed using any of the sophisticated methods available, e.g. RANSAC [8], the automatic homography estimation algorithm from [6].

3 Results

This section gives results of the performance of the proposed method. In a comparison with the automatic estimation algorithm from Hartley and Zisserman [6] section 4.8 it is shown how well the proposed algorithm performs in the case of low contrast infra red imagery. For the experiment a small part of a video sequence is used, consisting of 2201 frames. Throughout the video it can be seen that the camera performs small roll, pitch, and yaw movements while approaching the target area.

The proposed method is compared against the automatic homog-

raphy estimation algorithm. This algorithm uses interest points to estimate the specific homographies. These interest points are being matched between two consecutive frames of the video using a correlation matching technique. After these two steps point correspondences are given among two frames of the video. Using the corresponding points the RANSAC algorithm [8] is used to estimate a homography as initial value for an iterated optimization and guided matching step. The Levenberg-Marquardt algorithm [6] is used to perform the optimization step, by minimizing a symmetric cost function (9.2), over all point correspondences that were marked as inliers from the RANSAC algorithm.

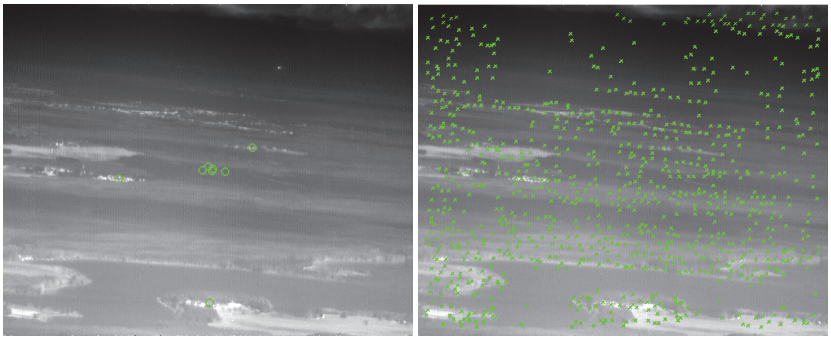
$$\sum_i d(x_i, \hat{x}_i)^2 + d(x_i\prime, \hat{x}_i\prime)^2 \quad \text{subject to } \hat{x}_i\prime = \hat{H}\hat{x}_i \; \forall i \tag{9.2}$$

The guided matching step uses the homography returned from the optimization step to search for new point correspondences among all point correspondences that were given as input to the RANSAC algorithm. As long as the number of point correspondences alters the optimization step and the guided matching step is iterated all over again. After this procedure the homographies for the video from the automatic algorithm are given.

The computation of the homographies based on the proposed method is done using the RANSAC algorithm. As stated in the previous section there is no need to do a matching step in order to compute corresponding points in two consecutive images. This is due to the fact that by using the specific camera matrices for each frame of the video the same GPS points can be projected into the specific frame and so the corresponding points between the consecutive frames are given naturally. Once the point correspondences are identified between all consecutive frames of the video, the RANSAC algorithm can be used to perform the homography estimation.

In order to show the different features on which both procedures are relying, figure 9.2 shows the GPS reference points (a) and the interest points (b) in the first image of the video sequence used.

Besides the distinction in establishing the homographies from frame to frame of the whole video sequence, a qualitative difference between both resulting homographies has to be evaluated. In order to do this, a region of interest is drawn into the first frame of the video sequence

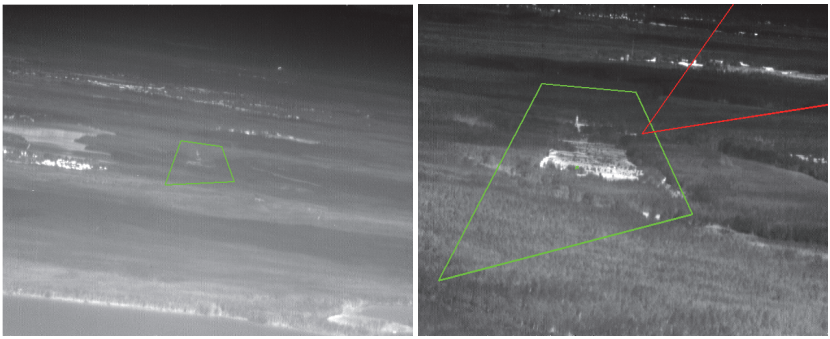


(a) (b)

Figure 9.2: (a) Projected GPS reference points in the first image of the video, (b) Interest points from the Foerstner operator in the first frame of the video.

for which the homographies were computed for. The region of interest in the first frame can be seen in figure 9.3 (a). Figure 9.3 (b) shows the last frame of the video with the transformed regions of interest of both methods. The region of interest of the proposed method is marked green and the one of the sophisticated method is marked red. The computed projective transforms of both methods are used to transform the regions of interest from frame to frame until the last image of the video is reached. While performing the transformations it was observed that the homographies of the proposed method performed well. There were trembling movements but no drifting, so the region of interest was staying on the designated target area. Whereas the homographies of the sophisticated method did not perform well. It was recorded that the region of interest began to waver and after a few frames started to drift towards the upper right corner of the image resulting in a position outside the scope of the last video frame.

When regarding the region of interest in the first frame, figure 9.3 (a), and the resulting regions of interest, figure 9.3 (b), it can be seen that the region of interest of the proposed method is framing the designated target area. In order to do some more sophisticated comparisons the positions of the corner points for both methods are given in table 9.2, together with the positions of the barycentre of the regions of interest as



(a) (b)

Figure 9.3: (a) Region of interest in the first frame of the video, (b) Resulting region of interest with homographies of proposed method (green) and of sophisticated procedure (red).

well as the size of the area enclosed by the regions of interest.

Looking at the results in table 9.2 it can be seen that the positions of the corner points of the proposed method are well proportioned compared to the resulting corner points of the sophisticated method. Also the enclosed area of both procedures shows that the proposed method behaved very well. This shows that the proposed method performs well even if the video material is quite noisy. Due to the problem of finding point correspondences in noisy image data it is inevitable to exploit additional knowledge about the imaging system.

4 Discussion

In this paper a method for the computation of homographies for videos with noisy image data is presented. The method exploits camera parameters and the position of the camera in world coordinates. The position in WGS84 coordinates is used to rotate the coordinates into the local coordinate system. Afterwards the roll, pitch, and yaw angles are used to transform the coordinates into the camera coordinate system. The focal length of the camera is used to form the camera projection matrix. The projection matrix of the camera is used to project GPS points into the

	frame $= 1$	**frame** $= 2200$	
		proposed	**sophisticated**
A	$\begin{pmatrix} 275 \\ 211 \end{pmatrix}$	$\begin{pmatrix} 192.23 \\ 123.68 \end{pmatrix}$	$\begin{pmatrix} 513.41 \\ -36.27 \end{pmatrix}$
B	$\begin{pmatrix} 338 \\ 220 \end{pmatrix}$	$\begin{pmatrix} 336.96 \\ 136.82 \end{pmatrix}$	$\begin{pmatrix} 1172.18 \\ -379.34 \end{pmatrix}$
C	$\begin{pmatrix} 357 \\ 273 \end{pmatrix}$	$\begin{pmatrix} 423.56 \\ 324.82 \end{pmatrix}$	$\begin{pmatrix} 1367.60 \\ 40.32 \end{pmatrix}$
D	$\begin{pmatrix} 251 \\ 278 \end{pmatrix}$	$\begin{pmatrix} 33.03 \\ 426.89 \end{pmatrix}$	$\begin{pmatrix} 346.87 \\ 200.36 \end{pmatrix}$
barycentre	$\begin{pmatrix} 305.25 \\ 245.5 \end{pmatrix}$	$\begin{pmatrix} 246.44 \\ 253.05 \end{pmatrix}$	$\begin{pmatrix} 850.02 \\ -43.73 \end{pmatrix}$
area	$5075\ \text{px}^2$	$\approx 64117\ \text{px}^2$	$\approx 297190\ \text{px}^2$

Table 9.2: Results from the comparison of the proposed method and the sophisticated method.

image plane. Together with the projected GPS points the homographies for the video sequence can be computed.

The results show that the proposed method works well using a RANSAC estimator for the computation of the homographies. The use of a more sophisticated estimator might improve the quality of the homographies e.g. the procedure of the automatic homography estimation algorithm without the feature matching step. The good performance of the proposed method is based on the fact that the projected image points are not affected by image noise, and the point correspondences are naturally given by the projection for the following frame. So the proposed method does not rely on feature matching methods to gather point correspondences.

In the case of noisy image data it is quite difficult to estimate homographies as can be seen in section 3. The introduction of a calibrated camera and knowledge about the position of the camera system can help to improve the estimated homographies.

References

1. H. P. Morevec, "Towards automatic visual obstacle avoidance," *Proceedings of the 5th International Joint Conference on Aritificial Intelligence*, vol. 2, pp. 584–584, 1977.
2. C. Harris and M. Stephens, "A combined corner and edge detector," *In Proc. of Fourth Alvey Vision Conference*, pp. 147–151, 1988.
3. M. A. Föstner and E. Gülch, "A fast operator for detection and precise location of distinct points, corners and centers of circular features," *ISPRS Intercommission Workshop*, 1987.
4. H. Sager, *Fourier-Transformation : Beispiele, Aufgaben, Anwendungen*, 1st ed., ser. vdf Lehrbuch Mathematik. Zürich: vdf Hochschulverl., 2012. [Online]. Available: http://www.vdf.ethz.ch/info/showDetails.asp?isbnNr=3393
5. C. K. Chui, *An introduction to wavelets*, ser. Wavelet analysis and its applications ; 1. San Diego [u.a.]: Academic Press, 1999.
6. R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.
7. NIMA, *Department of Defence World Geodetic System 1984*, 3rd ed., 2000. [Online]. Available: http://earth-info.nga.mil/GandG/publications/tr8350.2/wgs84fin.pdf
8. M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981. [Online]. Available: http://doi.acm.org/10.1145/358669.358692

Parameter-learning for color sorting of bulk materials using genetic algorithms

Matthias Richter[1] and Jürgen Beyerer[1,2]

[1] Karlsruhe Institute of Technology, Institute for Anthropomatics and Robotics, Adenauerring 4, D-76131 Karlsruhe
[2] Fraunhofer Insitute of Optronics, System Technologies and Image Exploitation (IOSB), Fraunhoferstr. 1, D-76131 Karlsruhe

Abstract Sensor based sorting finds broad applications in mining, recycling and quality control. Digital image processing and pattern recognition are key components, as they determine whether to keep or discard an object under inspection. In many scenarios, the color of a material stands out as the primary sorting criterion. In this paper, we present a flexible system for color sorting of bulk materials based on semantic color features. The features are constructed in a three stages: the color occurrence frequencies of different materials are estimated and then fused to a small number of color classes, which in turn are used to map each color to a discrete attribute. A compact object descriptor composed of the fractions of foreground pixels that share the same attribute characterizes the objects under inspection. This descriptor has many advantages: it has a very clear, intuitive interpretation, is invariant to rotation and scale of the object and requires very little computation. However, a major drawback are the many variables that govern the construction process. Manual fine tuning requires a large amount of time and experience. Subtle changes in the parameters can have strong effects on the classification performance. To overcome this shortcoming, we propose a method to automatically learn the parameters by a genetic algorithm. We apply our method to wine grape sorting problems to show that this approach performs at least as good a human expert. At the same time, it takes considerably less effort on the human part and frees the operator to attend to other tasks.

1 Introduction

The applications of optical sensor based sorting range from mining of precious metals and minerals over recycling of synthetics to quality control of food stuffs. Digital image processing and subsequent pattern recognition are key components, as in this stage it is determined whether to keep or discard the objects under inspection [1]. The research community has developed a multitude of approaches leveraging methods from computer vision. Without going into too much detail[3], these approaches usually involve extraction of low to mid-level features such as hue histograms, Gabor-descriptors and shape models, which are then fed into machine learning algorithms to derive a sorting decision.

However, such methods are rarely found in commercially available systems. There are two main reasons for this: Firstly, their black box nature prevents the operators to change classification parameters when the sorting requirements change. Secondly, both the feature extraction and classification algorithm often require extensive computation, which is infeasible considering the run time requirements of automated visual inspection.

Instead, commercial systems often combine several simple rules that put thresholds on simple features. Scott, for example, sorts plastic waste into two fractions by measuring the ratio of absorbances at two wavelengths in the infrared spectrum and comparing it to a threshold [4]. Similarly, Lee and Anbalagan put multiple thresholds on the red, green and blue color channels, which result in multiple accept and reject zones in the color space [5]. An object is kept if the color of its foreground-pixels fall mostly into accept zones, otherwise it is discarded. More recently, Blasco et al. presented their system to sort pomegranate arils [6]. Instead of directly defining decision regions in the RGB color space, they use a single threshold on the average ratio between the red and green color channel. This simple approach performed comparably to LDA using the whole RGB tuple as feature vector. The advantage of these approaches is that they can be implemented in hardware, which enables very high sorting speeds. Furthermore, the system's sorting criteria can easily be adjusted by changing the thresholds. On the other

[3] A full review is out of the scope of this paper. Interested readers are instead referred to the encompassing surveys by Du and Sun [2] as well as Malamas et al. [3].

hand, initial investigation to find suitable features and thresholds is a laborious process that has to be carried out by a trained expert.

Hybrid approaches apply thresholds to high level features that are learned from the color distribution of the materials under inspection. This combines the easy set-up of machine-learning approaches with the flexibility of commercial solutions. Duffy et al. detect burn marks on air filters by estimating the probability whether a pixel shows a defect [7]. They derive a histogram that characterizes the color of burn marks by building the difference of RGB histograms of intact and defective samples. Defects are located by back-projection and thresholding the resulting image. In a follow-up publication, Bergasa et al. do not estimate histograms directly but instead model the color distribution of defects as a mixture of Gaussians [8]. While slower in training, this approach proves more robust in testing, as it accounts for underrepresented and unseen defect appearances. Explicit modeling the color-distribution is not always needed. For inspection of color tablets in pharmaceutical blisters, Derganc et al. find optimal decision boundaries in the color space by employing a mode seeking algorithm [9]. In training, an operator marks a pixel belonging to a tablet. Then, a labeling function is constructed by determining the corresponding mode and subsequent cluster growing.

Applications are not limited to classification though. Lee et al. grade the maturity of dates according to the color of the fruit's surface [10]. In their analysis, they found that the color of dates of different ripeness fall into a thin connected region in the RGB space. Consequently, they find a projection onto a one-dimensional manifold by solving a second order trivariate polynomial regression problem. In a later publication, Zhang, Lee et al. simplified the process by estimating a back-projection table mapping RG-values to a ripeness level [11]. The table is built by first collecting characteristic RG-histograms of four maturity grades and fusing the histograms into a single lookup-table. Missing entries are filled in by linear interpolation using neighboring values.

While all these approaches show good results with their respective product, application to different problem domains is questionable. The methods presented by Duffy et al. [7,8] leave the question how to handle multiple defect classes. The mode finding approach by Derganc et al. [9] works well when the surface of objects under inspection is relatively uniform in color, but may struggle when the objects' color distribution is multimodal. Lee's method [10] requires the color distribution of each

class to be (approximately) supported on a one-dimensional connected manifold, which is seldom the case. The back-projection approach by Zhang et al. [11] alleviate this issue, but the construction of the lookup table is strictly tailored to grading date maturity and not directly applicable to other tasks.

With this in mind, we present a system capable of handling a large variety of products in different settings. The method is similar to the back-projection approaches presented above, but more general and not tailored to a specific product[4]. In a two-step process, we merge color histograms of different material fractions that may occur in the sorting application to color classes, which are then fused to build the lookup-table that maps RGB tuples to a discrete attribute. Objects are classified based on the fractions of pixels that map to each attribute.

2 Methods

In this section, we first describe our classification system and then turn our attention automatically learning the parameters that govern its behavior.

2.1 Classification System

Figure 10.1 shows the classification pipeline of our system. An input image $\mathcal{I}$ is transformed into an attribute image using a back-projection table that maps each RGB tuple $\mathbf{c} = (r, g, b)$ to an attribute,

$$\mathcal{A}(\mathbf{c}) = a \in \{-1, 0, 1, \ldots, M\}\,. \tag{10.1}$$

Here we use the convention that the attribute $\mathcal{A}(\mathbf{c}) = 0$ signifies a background pixel, whereas $\mathcal{A}(\mathbf{c}) = -1$ denotes an unknown color. Using blob analysis on the attribute image, single objects are extracted. A feature vector $\mathbf{m} = (m_{-1}, m_1, \ldots, m_M)^\top$ is calculated for each object, where each entry m_i in $\mathbf{m}$ represents the fraction of pixels that map to the i-th attribute. Formally, with $\mathcal{O}$ denoting the set of foreground pixels:

$$m_i = \frac{1}{|\mathcal{O}|} \sum_{(x,y)\in\mathcal{O}} \mathbb{1}[\mathcal{A}(\mathcal{I}(x, y)) = i]\,. \tag{10.2}$$

[4] In fact, we applied our method in different scenarios from recycling to food-inspection.

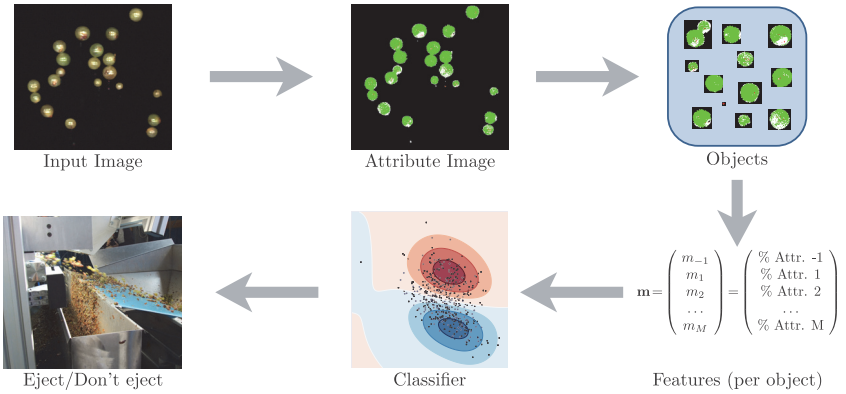


Figure 10.1: Overview of the classification pipeline.

The decision whether to keep or discard an object is done by a binary classifier $H(\mathbf{m}) = y \in \{-1, 1\}$ and a signal is sent to the actuating hardware. The choice of classifier is arbitrary, however keeping the intended application in mind, we typically settle for simple rule-based classifiers.

Color Attributes

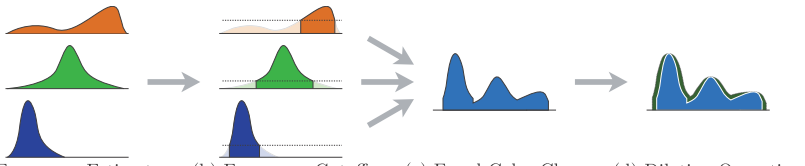
Key to this classification pipeline is the mapping from color to attribute in eq. (10.1). Figure 10.2 outlines the steps performed to derive color classes, which are the basic building blocks of the attribute-mapping $\mathcal{A}(\mathbf{c})$. In detail, the process is as follows:

The materials expected to be encountered during sorting are placed into the sorting machine and images are captured as if the system was in operation. From these images, color frequency estimates

$$\hat{p}_\kappa(\mathbf{c}|k), \quad k = 0, 1, \ldots, K \tag{10.3}$$

are collected. As with the attribute-mapping in eq. (10.1), the index $k = 0$ denotes the background. The remaining $\hat{p}_\kappa(\mathbf{c}|k)$ are estimated only from foreground pixels. The choice of estimator is arbitrary, but for the sake of simplicity we chose to use the joint RGB histogram.

If the ground-truth images show dirt particles or other non-target materials, the histograms may contain non-informative entries, which



(a) Frequency Estimates (b) Frequency Cutoff (c) Fused Color Class (d) Dilation Operation

Figure 10.2: Outline of the steps performed to derive color classes that are fused into the attribute table. Note that the figure shows 1D-histograms, while our system is based on dense, 3D-RGB histograms.

can be disruptive in later stages of the pipeline. Addressing this issue, we drop frequencies below a user-definable threshold $\beta \in [0, 1]$ (see Fig. 10.2 (b)). Formally, this amounts to the following operation:

$$\mathrm{cut}(\hat{p}_\kappa(\mathbf{c}|k), \beta) = \begin{cases} \frac{1}{Z}\,\hat{p}_\kappa(\mathbf{c}|k) & \text{if } \hat{p}_\kappa(\mathbf{c}|k) \geq \beta \\ 0 & \text{otherwise.} \end{cases} \tag{10.4}$$

Here Z is a normalization constant so that $\mathrm{cut}(\hat{p}_\kappa(\mathbf{c}|k), \beta)$ is a probability distribution. The resulting modified frequency estimates are fused into color classes (see Fig. 10.2 (c)) by weighted averaging,

$$\hat{p}_\mu(\mathbf{c}|m) = \sum_{k=0}^{K} \alpha_{km}\ \mathrm{cut}(\hat{p}_\kappa(\mathbf{c}|k), \beta_k)\,, \quad m = 0, 1, \ldots, M, \tag{10.5}$$

where $\alpha_{km} \geq 0$ and $\sum_k \alpha_{km} = 1$. The color classes $\hat{p}_\mu(\mathbf{c}|m)$ represent higher-level features and correspond to semantic groups of materials.

The color frequency estimates and therefore the color classes are generally derived using a relatively small sample. As a consequence, the estimates may not accurately reflect the underlying distribution, especially when considering natural materials, whose appearance can fluctuate in time. Furthermore, external influences such as stray light can additionally alter the perceived color of the objects. To alleviate these problems, we apply a 3D morphological filter on the color classes:

$$\mathrm{dilate}(\hat{p}_\mu(\mathbf{c}|m), \delta) = \max_{\mathbf{o} \in [-\delta, \delta]^3} \hat{p}_\mu(\mathbf{c} - \mathbf{o}|m). \tag{10.6}$$

Finally, the attribute table is constructed by assigning the color class

with the highest weighted probability for the given color,

$$\mathcal{A}(\mathbf{c}) = \begin{cases} -1 & \text{if } \max_m\{D_m\} = 0 \\ \arg\max_m \{\gamma_m D_m\} & \text{else,} \end{cases} \tag{10.7}$$

where $D_m = \text{dilate}(\hat{p}_\mu(\mathbf{c}|m), \delta_m)$ and $\gamma_m \geq 0$. Incidentally, equation (10.7) can be interpreted as a maximum a posteriori classifier that assigns pixels to a color class, where γ_m encodes the class prior.

2.2 Parameter Learning

As mentioned in the previous section, the attribute table has a significant impact on the overall classification performance, more so than the classifier itself. However, finding the optimal configuration is a time consuming process where seemingly small adjustments can cause notably different sorting results. We therefore propose to automatically estimate good parameter combinations. Ideally, the user should only provide the initial color frequency estimates and the desired number of color classes M and the computer should figure out the remaining parameters.

One way to achieve this goal is to pose the task as optimization problem, i.e. to search the set of parameters that achieves the optimal classification performance.

Genetic Algorithms

Genetic algorithms (GA) are a well-known meta-heuristic to find a set of parameters $\boldsymbol{\theta}$ that maximize a fitness (or merit) function $f(\boldsymbol{\theta})$. GAs have been shows to work well with large or even infinite search spaces and are able to find the global optimum in non-convex problems [12]. However, the solution is generally only approximately optimal: With $\boldsymbol{\theta}^\star$ denoting the true optimum and $\varepsilon > 0$, a GA finds a solution $\hat{\boldsymbol{\theta}}$ with

$$|f(\boldsymbol{\theta}^\star) - f(\hat{\boldsymbol{\theta}})| < \varepsilon. \tag{10.8}$$

The method is modeled after the theory of natural selection: A population of individuals (possible solutions) produce offspring (new solution candidates) through recombination. The offspring is subject to

random mutation and the fittest individuals are selected according to the fitness function f. The process repeats until a certain number of iterations is reached. The following pseudo-code outlines the approach:

Input: Population size N, Number of iterations K
Output: Candidate solutions $\mathcal{P}$

Randomly sample population $\mathcal{P} = \{\boldsymbol{\theta}_n | n = 1, \ldots, N\}$
for $k = 1$ **to** K:
    $\mathcal{C} \leftarrow \mathrm{crossover}(\mathcal{P})$
    $\mathcal{C} \leftarrow \mathrm{mutate}(\mathcal{C})$
    $\mathcal{P} \leftarrow \text{select-fittest}(\mathcal{P} \cup \mathcal{C})$
return $\mathcal{P}$

Key components are $\mathrm{crossover}$ and mutate operations, which explore the parameter-space around the existing solutions. There exists several alternatives to perform $\mathrm{crossover}$, but here we focus on tournament selection with random recombination. In random recombination, the offspring $\boldsymbol{\theta}_c$ of two parent individuals $\boldsymbol{\theta}_p$ and $\boldsymbol{\theta}_q$ is produced by randomly choosing $\theta_c^{(i)}$ as the i-th element of either parent with equal probability. Each parent is selected by randomly sampling two candidates from the population and keeping the fitter one, i.e. for $\boldsymbol{\theta}_p$:

$$\boldsymbol{\theta}_p = \arg\max\{f(\boldsymbol{\theta}_{p_1}), f(\boldsymbol{\theta}_{p_2})\}, \quad \boldsymbol{\theta}_{p_1}, \boldsymbol{\theta}_{p_2} \in \mathcal{P}. \tag{10.9}$$

Since crossover alone is not sufficient to fully explore the parameter space, mutate performs a randomized local search by randomly changing elements of each $\boldsymbol{\theta}_c \in \mathcal{C}$. Finally select-fittest ranks all $\boldsymbol{\theta} \in \mathcal{P} \cup \mathcal{C}$ according to their fitness $f(\boldsymbol{\theta})$ and keeps only the N best-performing individuals.

The repeated application of crossover, mutation and selection have the effect that the population, which is initially scattered around the parameter space, converges onto a maximum of the fitness function. Due to the inherent randomization, GAs can recover from falling into local minima, which sets them apart from other methods such as gradient descent and hill-climbing. Furthermore, constraints on the parameters are almost trivial to implement by adjusting the $\mathrm{crossover}$ and mutate operations and regularization is achieved by adding an appropriate term to the fitness function.

Application

Due to the non-convex nature of our parameter-optimization and the large search space, we chose GAs to find good combinations of parameters. We constrain the parameters according to Section 2.1, i.e.

$$\alpha_{km}, \gamma_m \geq 0, \text{ with } \textstyle\sum_k \alpha_{km} = 1, \tag{10.10}$$

$$0 \leq \beta_k \leq 1 \text{ and} \tag{10.11}$$

$$\delta_m \in \mathbb{N}_0. \tag{10.12}$$

As fitness function we employ Matthews correlation coefficient [13], which can be interpreted as the correlation coefficient between ground-truth and prediction of the classifier $H(\mathbf{m})$. With n_{tp}, n_{fp}, n_{tn} and n_{fn} denoting the number of true positive, false positive, true negative and false negative classifications on a validation sample it can be defined as

$$\text{MCC} = \frac{n_{\text{tp}}\, n_{\text{tn}} - n_{\text{fp}}\, n_{\text{fn}}}{\sqrt{(n_{\text{tp}} + n_{\text{fp}})(n_{\text{tp}} + n_{\text{fn}})(n_{\text{tn}} + n_{\text{fp}})(n_{\text{tn}} + n_{\text{fn}})}}. \tag{10.13}$$

3 Experiments

We validated our approach by comparing the classification performance of hand tuned parameters to parameters learnt by the GA. We considered the sorting problem of discriminating healthy wine berries from grapes with fungal infection and unwanted parts of the plant. Experiments were performed with three varieties of wine: Pinot noir, Pinot blanc and Riesling. All images were acquired using an off-the-shelve RGB line-scan camera. Since berries of the Pinot noir variety are very dark and show low contrast to the black background, the blue channel was replaced with a NIR-channel in this case. To reduce the parameter space, we do not perform frequency thresholding (i.e. set $\beta_k = 0$ for all k) and constrain the background class to not include, and not to be included in the other color classes by setting $\alpha_{00} = 1$ and $\alpha_{0m} = \alpha_{k0} = 0$. Parameters were mutated by each selecting a random α_{km}, γ_m and δ_m and assigning a new value in the variable's domain with probability of $p = 0.8$, $p = 0.5$ and $p = 0.3$ respectively. We chose a linear SVM as classifier $H(\mathbf{m})$, since it is relatively fast to train and evaluate, and therefore reduces the time required for the optimization.

Grape variety	# of samples		MCC wrt. selection method	
	positive	negative	manual	learned
Pinot noir	2416	641	0.47	0.70 $\pm$0.12
Pinot blanc	332	291	0.86	0.86 $\pm$0.17
Riesling	1061	235	0.88	0.86 $\pm$0.11

Table 10.1: Overview over the results of our experiments.

Table 10.1 shows the results of our experiments. In case of learned parameters, we performed stratified 10-fold cross-validation to estimate the mean and standard deviation of MCC values. We did not perform cross-validation with hand-tuned parameters. In each fold, the training set was randomly split in two subsets, where the first one was used to estimate parameters and the second was used to train the classifier. Especially with the Pinot blanc variety, this resulted in very few training samples and subsequently relatively high variance of the classification performance. Nonetheless, the classification performance is on par with manually estimated parameters. In case of Pinot noir, the GA approach even outperformed the human expert, who had difficulties finding an appropriate set of parameters due to the (apparent) lack of an informative blue channel. Table 10.2 shows the learned parameters of the best performing solution with the Pinot noir variety. While the first color class puts more emphasis on the third and fourth color frequency estimate (defects), the second color class is influenced primarily by the fifth and sixth frequency estimate (ripe berries). This roughly corresponds to $m = 1$ and $m = 2$ denoting *defect*- and *accept*-classes. However, both classes consider all available color estimates (apart from the background – which was enforced by the experimental constraints). This contrasts with the approach of a human expert, who typically selects few color frequency estimates to build the color classes.

4 Conclusion

In this paper we have presented both a method for color classification of bulk materials and a method to automatically estimate the parameters governing the classification. The system is flexible and can accommo-

α_{km}	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	γ_m	δ_m
$m=0$	1.00	0.00	0.00	0.00	0.00	0.00	0.06	1
$m=1$	0.00	0.16	0.27	0.47	0.09	0.01	1.00	3
$m=2$	0.00	0.07	0.01	0.18	0.41	0.34	0.08	3

Table 10.2: Best performing learned configuration for the Pinot blanc variety.

date a wide range of materials. While manual set-up is a labor intensive and time-consuming process, parameter learning only requires human interaction when estimating the basic color distributions. The remaining learning process is fully automatic and frees the human to attend to other tasks. Moreover, it requires no knowledge of the inner workings of the systems and therefore allows non-experts to bootstrap a working classifier.

However, there is still room for improvement. In particular, one could use kernel density estimators instead of joint RGB histograms to derive the color frequencies $\hat{p}_\kappa(\mathbf{c}|k)$. Doing so would remove the need to perform dilation (10.6) and therefore reduce the dimensionality of the parameter space. Another modification would be to encourage sparsity of the α_{km} by including an appropriate term in the fitness function. The resulting parameters would be more similar to configuration set up by a human expert and thus be more open to interpretation. Lastly, the genetic algorithm could be replaced by other heuristic optimization procedures such as particle swam optimization, which puts a stronger emphasis on local search, or simulated annealing, which evaluates the merit function less frequently than a GA and thus might be faster to find the optimum.

References

1. H. Wotruba and H. Harbeck, "Sensor-based sorting," *Ullmann's Encyclopedia of Industrial Chemistry*, 2010.

2. C.-J. Du and D.-W. Sun, "Learning techniques used in computer vision for food quality evaluation: a review," *Journal of Food Engineering*, vol. 72, no. 1, pp. 39–55, Jan. 2006.

3. E. N. Malamas, E. G. Petrakis, M. Zervakis, L. Petit, and J.-D. Legat, "A survey on industrial vision systems, applications and tools," *Image and Vision Computing*, vol. 21, no. 2, pp. 171–188, Feb. 2003.
4. D. M. Scott, "A two-colour near-infrared sensor for sorting recycled plastic waste," *Measurement Science and Technology*, vol. 6, no. 2, pp. 156–159, Feb. 1995.
5. D.-J. Lee and R. S. Anbalagan, "High-speed automated color-sorting vision system," in *Optical Engineering Midwest*, vol. 2622. International Society for Optics and Photonics, Aug. 1995, pp. 573–579.
6. J. Blasco, S. Cubero, J. Gómez-Sanchís, P. Mira, and E. Moltó, "Development of a machine for the automatic sorting of pomegranate (Punica granatum) arils based on computer vision," *Journal of Food Engineering*, vol. 90, no. 1, pp. 27–34, Jan. 2009.
7. N. Duffy, J. Crowley, and G. Lacey, "Object detection using colour," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 1. IEEE Comput. Soc, 2000, pp. 700–703.
8. L. Bergasa, N. Duffy, G. Lacey, and M. Mazo, "Industrial inspection using Gaussian functions in a colour space," *Image and Vision Computing*, vol. 18, no. 12, pp. 951–957, Sep. 2000.
9. J. Derganc, B. Likar, R. Bernard, D. Tomaževič, and F. Pernuš, "Real-time automated visual inspection of color tablets in pharmaceutical blisters," *Real-Time Imaging*, vol. 9, no. 2, pp. 113–124, Apr. 2003.
10. D.-J. Lee, J. K. Archibald, Y.-C. Chang, and C. R. Greco, "Robust color space conversion and color distribution analysis techniques for date maturity evaluation," *Journal of Food Engineering*, vol. 88, no. 3, pp. 364–372, 2008.
11. D. Zhang, D.-J. Lee, B. J. Tippetts, and K. D. Lillywhite, "Date maturity and quality evaluation using color distribution analysis and back projection," *Journal of Food Engineering*, vol. 131, pp. 161–169, Jun. 2014.
12. M. Mitchell, *An introduction to genetic algorithms*, 1998.
13. B. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochimica et Biophysica Acta (BBA) - Protein Structure*, vol. 405, no. 2, pp. 442–451, Oct. 1975.

Mehrdimensionale Merkmale zur Augendetektion

Sebastian Vater und Fernando Puente León

Karlsruher Institut für Technologie,
Institut für Industrielle Informationstechnik,
Hertzstraße 16, 76187 Karlsruhe, vater@kit.edu, +49 721 608-44518

Zusammenfassung In diesem Beitrag wird eine Methode zur Kombination mehrdimensionaler, erscheinungsbasierter Merkmale mit einem Kaskadenklassifizierer zur Augendetektion vorgestellt. Dem Problem der Hochdimensionalität wird mit einem Boosting-Ansatz zur Reduktion der Merkmalsdimensionen begegnet. Durch *Merkmalsbagging* ist das Verfahren in der Lage, verschiedene Merkmalstypen effizient in einem Klassifizierer zu kombinieren.

1 Einleitung

Ein wichtiger Bestandteil der Mensch-Maschine-Interaktion ist das Erkennen sowie Lokalisieren von Objekten der Klasse Auge in einer unbekannten Umgebung. Aus dem Bereich der Augen lassen sich aus Bildsequenzen wichtige Informationen über die Aufmerksamkeit oder die Blickrichtung erschließen. Um eine Verwendung mit einfacher Hardware, wie etwa Webcams oder Smartphones, zu ermöglichen, sind robuste, zuverlässige Methoden sowie Echtzeitfähigkeit notwendig.

Bei der Augendetektion stellt die große Variation innerhalb der Klasse Auge sowie – bedingt durch eine geringe Auflösung – der oftmals kleine Bildbereich, welcher das Auge beschreibt und somit nur eine begrenzte Menge an Information bereitstellt, eine Herausforderung dar, die eine Erweiterung bestehender Ansätze für die Objektdetektion in Echtzeit erfordert. Eine Verbesserung der Robustheit ist insbesondere für starke Ausprägungen des Blickwinkels erforderlich. Existierende Methoden teilen das gemeinsame Problem der Trennbarkeit der Klassen bei hoher Varianz der Trainingsdaten. Dies führt zur Einschränkung

der Konvergenz des Trainings, während eine detailliertere Beschreibung des Bildinhalts durch höherdimensionale Merkmale den Rechenaufwand erhöht. Die in dieser Arbeit vorgestellte Methode nimmt sich dieses Problems an, indem folgende Beiträge geliefert werden: Zum einen wird das Training eines Kaskadenklassifizieres mit verschiedenen Merkmalstypen durch kaskadiertes Boosting vorgestellt. Zweitens wird eine effiziente Methode zur Einbindung mehrdimensionaler Merkmale in den Algorithmus präsentiert. Den dritten Beitrag stellt ein eigens zusammengesetzer Trainingsdatensatz dar, der ein bezüglich des Kopfkoordinatensystems örtlich stationäres Detektionsfenster liefert, dessen Mittelpunkt als Approximation der Irisposition dient.

2 Stand der Wissenschaft

Bestehende Verfahren verwenden meist erscheinungsbasierte Ansätze [1], bei denen Merkmale in einem Boosting-Algorithmus effizient eingesetzt werden [2]. Hochdimensionale Merkmale werden auf Kosten des Rechenaufwandes zusammen mit Support-Vector-Machines (SVM) eingesetzt, um die Robustheit zu erhöhen [3]. Verbesserungen werden durch Kombination von zwei Merkmalstypen [4] oder durch Integration einer SVM in einen Boosting-Algorithmus erzielt [5]. Forschungsbedarf resultiert aus einer unzureichenden Performanz [6] sowie auf Grund des Fehlens einer Validierung der Robustheit der Verfahren unter Berücksichtigung einer veränderlicher Blickrichtung.

3 Kaskadenklassifizierer

Das Rahmenwerk dieses Beitrags liefert eine Implementierung des in [2] vorgestellten Kaskadenklassifizierers. Abbildung 11.1 zeigt schematisch den Aufbau des implementieren Klassifizierers.

3.1 Detektion

Bestehend aus N starken Klassifikatoren K wird jedes Teilfenster eines Eingangsbildes $g(\mathbf{x})$, welches durch seine Position p_x, p_y sowie seine Höhe h und Breite b charakterisiert ist, separat klassifiziert. Dabei

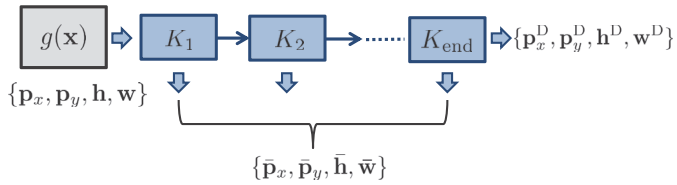


Abbildung 11.1: Schematische Darstellung des Kaskadenklassifizierers. Mit K_i werden die einzelnen Knoten, die jeder für sich einem Klassifizierer entsprechen, bezeichnet. Die Positionsvektoren der Kandidaten für Detektionsfenster sind mit $\mathbf{p}_x$, $\mathbf{p}_y$ beschrieben. Ein Querbalken bedeutet ein zurückgewiesenes Fenster; ein hochgestelltes D bestätigt eine Detektion an der entsprechenden Position.

soll jeder Knoten K_i lediglich eine geforderte Detektionsrate sowie eine maximal zulässige Falschalarmrate erfüllen. Durch Serienschaltung der Knoten nimmt die Falschalarmrate mit der Potenz N ab. Eine Detektion erfolgt, wenn ein Teilfenster alle Knoten erfolgreich durchlaufen hat. Die Effizienz der Detektors ist durch die Zunahme der Komplexität der Klassifikationsentscheidung gegeben. Während zu Beginn noch alle Teilfenster zu klassifizieren sind, ist hier die Komplexität gering und die Entscheidung kann mit wenig Aufwand getroffen werden. Mit jeder Stufe nimmt die Anzahl der zu betrachtenden Teilfenster mit der Falschalarmrate (plus Anzahl korrekt detektierter Objekte) ab.

3.2 Training

Das Training des Klassifikators basiert auf dem in [7] vorgeschlagenen Boosting. Es werden hierzu die Antworten aller Merkmale $\mathbf{m}(\mathbf{x})$ aus einem Merkmalspool auf Trainingsbeispiele für die Klassen *Auge* und *Kein Auge* berechnet. Durch Boosting wird nun das diskriminativste Merkmal $m_t(\mathbf{x})$ ausgewählt und eine Gewichtung β_t mit Hilfe einer Verlustfunktion bestimmt, welche sich in Abhängigkeit des Klassifikationsfehlers e_t des gewählten Merkmals ergibt:

$$\beta_t = \frac{e_t}{1 - e_t} \,. \tag{11.1}$$

Die Gewichtung β_t der richtig klassifizierten Trainingsbeispiele erfolgt nach jeder Merkmalsauswahl. Dadurch soll nach einer Normalisierung der Gewichte sichergestellt werden, dass sich der Klassifizierer bei der fortlaufenden Merkmalsauswahl auf schwierig zu entscheidende Trainingsbeispiele konzentriert. Durch die Gewichtung mit dem negativen Logarithmus der Verlustfunktion

$$\alpha_t = -\log \beta_t \tag{11.2}$$

werden anschließend die Merkmale zu einem starken Klassifikator K geboostet:

$$K(\mathbf{x}) = \operatorname{sign}\left(\sum_t \alpha_t \, h_t(\mathbf{x}) - \tau\right) .$$

Hier stellt $h_t(\mathbf{x}) \in \{-1, 1\}$ die Hypothese des Merkmals $m_t(\mathbf{x})$ und τ einen Schwellwert dar.

4 Merkmale

Der Beitrag verfolgt den Ansatz, die Deskriptivität verschiedener Merkmale zu kombinieren, indem zunächst eine große Menge an Merkmalen erzeugt wird. Diese werden dann mittels Boosting durch eine Gewichtung einzelner schwacher Klassifikatoren zu einem starken Klassifikator verstärkt, welcher die Klassifikation optimiert.

4.1 Eindimensionale Merkmale

Ursprünglich wurden in [1] ausschließlich Haar-Wavelets erster Ordnung verwendet. Mittels Gleichung (11.3) können sie effizient unter Verwendung der integralen Bilddarstellung ausgewertet werden:

$$m_{\mathrm{Haar}}(\mathbf{x}) = \sum_j \frac{1}{|\mathcal{A}_j|} \sum_{\mathbf{u} \in \mathcal{A}_j(\mathbf{x})} w_j \, g(\mathbf{u}) \, . \tag{11.3}$$

Hierbei bezeichnen $g(\mathbf{x})$ das betrachtete Grauwertbild, $\mathbf{x} = (x, y)^{\mathrm{T}}$ den diskreten Ort, $\mathcal{A}_j(\mathbf{x})$ eine Umgebung des Ortes $\mathbf{x}$ mit der Fläche $|\mathcal{A}_j|$

und w_j einen Gewichtungsfaktor für die Umgebung $\mathcal{A}_j$. Während Gleichung (11.3) lokale Grauwertunterschiede beschreibt, lassen sich mit Gleichung (11.4) Gradienteninformationen erfassen. Sie sind auch unter dem Namen „*Edge Density*"-Merkmale (ED) [8] bekannt:

$$m_{\mathrm{ED}}(\mathbf{x}) = \frac{1}{|\mathcal{A}_j|} \sum_{\mathbf{u}\in\mathcal{A}_j(\mathbf{x})} |\Theta(\mathbf{u})| \,. \tag{11.4}$$

Eine Einschränkung der ED-Merkmale ist, dass lediglich der Gradientenbetrag genutzt wird und die Richtung vernachlässigt wird. *Edge of Oriented Histograms* (EOH) bieten eine effizient zu implementierende Möglichkeit, Richtungsinformation der Gradienten mit in die Merkmalsberechnung einzubeziehen. In Gleichung (11.5) stellt d die Anzahl der Richtungen, in welche die Gradienteninformation quantisiert wird, dar:

$$m_{\mathrm{EOH}}(\mathbf{x}) = \sum_{\mathbf{u}\in\mathcal{A}_j(\mathbf{x})} \frac{|\Theta(\mathbf{u})|\,\delta(\angle(\Theta(\mathbf{u}) - \gamma))}{|\Theta(\mathbf{u})|} \tag{11.5}$$

mit

$$\gamma \in [\,\psi_0,\, \psi_0 + \Delta\psi)\,, \quad \Delta\psi = \frac{2\pi}{d}, \quad d \in \mathbb{N}\,, \quad \psi_0 \in \mathbb{R}\,.$$

Eine in diesem Beitrag vorgestellte Erweiterung des EOH-Merkmals stellen die „*Neighbourhood Normalized EOH*"-Merkmale (NNEOH) dar. Im Unterschied zu den gewöhnlichen EOH-Merkmalen wird eine robustere Normalisierung gegenüber Beleuchtungsunterschieden durch eine großflächigere Normalisierung erreicht, wobei im Nenner in Gleichung (11.5) ein Bereich $|\mathcal{A}_j^{\mathrm{NNEOH}}| \geq |\mathcal{A}_j^{\mathrm{EOH}}|$ berücksichtigt wird.

Im folgenden Abschnitt wird eine Methode zur Dimensionsreduktion mehrdimensionaler Merkmale vorgestellt. Um die Robustheit des Detektors und die Deskriptivität der Merkmale weiter zu erhöhen, sollen hiermit „*Histogram of Oriented Gradients*"-Merkmale (HOG) in den hier vorgestellten Klassifikationsalgorithmus implementiert werden. Dabei soll die Deskriptivität der Merkmale erhalten bleiben und der Rechenaufwand der hochdimensionalen HOG-Merkmale verringert werden.

4.2 Merkmalsreduktion

Das Verfahren zur Merkmalsreduktion baut auf dem Boostingansatz auf. Nachdem durch Skalierung und Translation ein Pool aus blockweise normalisierten Merkmalen der Dimension d multipliziert mit der Anzahl an Blöcken definiert wurde, wird der eindimensionale Merkmalswert berechnet. Tabelle 11.1 beschreibt das Vorgehen zur Berechnung der dimensionsreduzierten HOG-Merkmale (RHOG) in Form von Pseudocode. Der große Vorteil des vorgestellten Verfahrens ist seine Generizität. Während in [3] eine aufwändige Suche nach der idealen Parametrierung der Zellen- und Blockgröße der HOG-Merkmale in Abhängigkeit der Trainingsdaten durchgeführt werden muss, werden hier die deskriptivsten Merkmale in Abhängigkeit von Position und Skalierung aus dem Merkmalspool bestimmt.

Tabelle 11.1: Pseudocode zur Beschreibung der Merkmalsreduktion.

- Für alle Dimensionen des Merkmals:
  - Berechne Merkmalsantworten aller Merkmale auf Datensatz
  - Finde minimalen Fehler und Schwellwert für Dimension
  - Bestimme Gewicht für Dimension mit Gleichungen (11.1) und (11.2)
  - Berechne Hypothesen für Merkmalsdimension
- Berechne 1-D Merkmalswert durch Hypothesen und Gewichte

4.3 Merkmalsbagging

Durch *Merkmalsbagging* wird dem Problem des mit der Verwendung vieler Merkmalstypen einhergehenden Bedarfs an Speicherplatz mit einem vorgeschalteten Boosting begegnet. Dazu wird zunächst sukzessiv nur mit aus einem Merkmalstyp bestehenden Bag trainiert und so eine Vorauswahl getroffen. Anschließend werden die in den einzelnen Trainingsschritten ausgewählten Merkmale in einem neuen Bag of Features $\hat{\mathbf{M}}$ zusammengefasst. Somit kann eine geeignete Vorauswahl der Merkmale für das finale Training realisiert werden. Abbildung 11.2 zeigt diesen Vorgang graphisch.

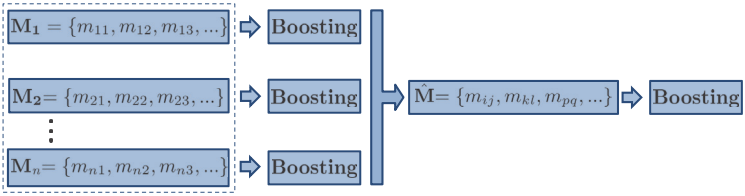


Abbildung 11.2: Schematische Darstellung der kaskadierten Merkmalsauswahl (*Merkmalsbagging*). Die Indizes ij, kl, pq sollen andeuten, dass sich die resultierende Merkmalsmenge aus beliebigen ursprünglichen Mengen zusammensetzen kann. $\mathbf{M_1}$, $\mathbf{M_2}$,..,$\mathbf{M_n}$ bezeichnen die Bags der verschiedenen Merkmalstypen.

5 Datensatz

Der Trainingsdatensatz besteht aus Ausschnitten der Augenregionen aus den Datensätzen *BioID* [9], *University of Essex Face Database* [10], *Utrecht Faces Database* [11] sowie *Yale Face Database B* [12]. Insgesamt wurden über 11000 positive Beispiele zum Training genutzt. Um das Detektionsfenster innerhalb des Suchbereiches um die Irisposition herum zu stabilisieren, wurden die zum Training genutzten Daten zentriert um das Auge herum ausgeschnitten. Dadurch wird sichergestellt, dass die ortsfesten Merkmale stabil auf lokale Charakteristiken reagieren und nicht etwa auf unpräzise Bereiche, wie die dunklen und hellen Regionen, durch sich grob die Augenbrauen beziehungsweise die Sclera kennzeichnen lassen.

6 Ergebnisse

Die Ergebnisse wurden anhand des *Caltech 101 Dataset* [13] mit 450 Bildern frontal aufgenommener Gesichter ausgewertet. Abbildung 11.3 zeigt die mit den beschrieben Verfahren erzielten Ergebnisse für die Detektionsraten in Abhängigkeit der Falschalarmrate. Es ist deutlich zu erkennen, dass eine Hinzunahme mehrerer Merkmalstypen eine Performanzsteigerung hervorruft. Das Diagramm zeigt einen Vergleich mit der „*haarcascade_mcs_eyepair_small*"-Kaskade, die Bestandteil von Matlab 2012a ist.

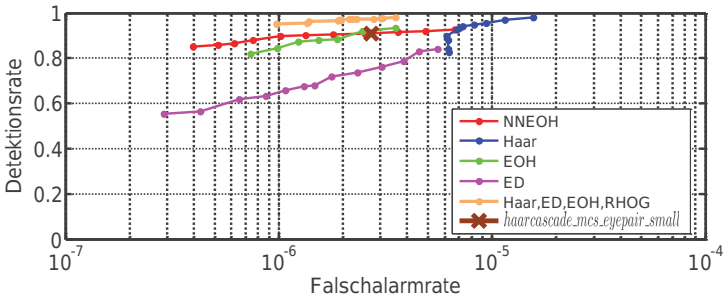


Abbildung 11.3: Performanz des Kaskadenklassifizierers durch Verwendung von *Merkmalsbagging* und dem Einbinden von mehrdimensionalen Merkmalen mit Hilfe der vorgestellten Merkmalsreduktion.

Abbildung 11.4 zeigt im rechten Bild eine Erklärung für den Performanzsprung bei zusätzlich verwendeten RHOG-Merkmalen. Trotz starker Beleuchtungsunterschiede ist die Detektion erfolgreich. Um eine Aussage über die Güte der Detektion bezüglich der Irisposition zu machen, wurde das in [14] vorgeschlagenen Fehlermaß verwendet:

$$d = \frac{\max(d_1, d_2)}{\Delta C} .$$

Hier beschreiben d_1, d_2 die Abweichungen zwischen Ground Truth Irismittelpunkt und dem Detektionsergebnis für das linke und rechte Auge basierend auf dem Detektionsfenster. Normiert wird das Fehlermaß mit dem Abstand der beiden Augen ΔC. Der Fehler in Abbildung 11.4 im linken Bild beträgt $d = 0{,}04$.

6.1 Zusammenfassung

Der vorgestellte Beitrag präsentiert ein Rahmenwerk zur effizienten Kombination ein- und mehrdimensionaler, erscheinungsbasierter Merkmale in einem Kaskadenklassifikator. Der Schwierigkeit der Augendetektion, die aufgrund ihrer großen Intraklassenvarianz gegeben ist, wird mit einem breiten Spektrum an Merkmalsdeskriptoren begegnet. Mit Hilfe kaskadierten Boostings kann dieser Merkmalspool effizi-

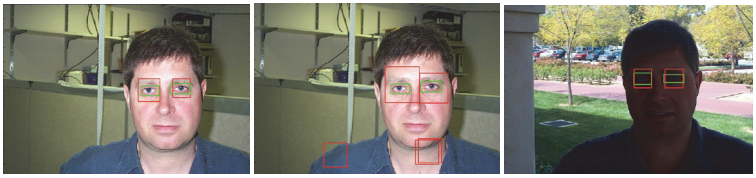


Abbildung 11.4: Detektionsergebnisse auf dem *Caltech 101 Dataset*. Links und Mitte: Vergleich der besten in dieser Arbeit erstellten Kaskade und einer Matlab Kaskade zur Augendetektion. Rechts: Detektion bei schwierigen Beleuchtungsbedingungen. In Grün sind Ground Truth Daten, in Rot Detektionsergebnisse skizziert.

ent verarbeitet werden. Es wird eine Methode vorgestellt, mit Hilfe derer hochdimensionale Merkmale, wie etwa HOG-Merkmale, unter Beibehaltung ihrer Deskriptivität bezüglich ihrer Dimension und dem damit verbundenen Rechenaufwand reduziert werden können.

Weiterführend soll ein am Institut für Industrielle Informationstechnik aufgenommener Datensatz bestehend aus Bildern von Augen mit stark variierenden Blickwinkeln in das Training integriert werden. Mit dem beschriebenen Verfahren sollen dann weiterhin Texturinformation durch LBP-Merkmale integriert werden und so die Deskriptivität und Robustheit des Detektors weiter verbessert werden. Eine umfassende Auswertung der detektierten Irispositionen soll folgen.

Literatur

1. M. Oren, C. Papageorgiou, P. Sinha, E. Osuna und T. Poggio, „Pedestrian detection using wavelet templates“, in *Computer Vision and Pattern Recognition, 1997. Proceedings, IEEE Computer Society Conference on*, S. 193–199.
2. P. Viola und M. Jones, „Rapid object detection using a boosted cascade of simple features“, in *Computer Vision and Pattern Recognition. CVPR 2001. Proceedings of the IEEE Computer Society Conference on*, Vol. 1, S. I–511–I–518.
3. N. Dalal und B. Triggs, „Histograms of oriented gradients for human detection“, in *Computer Vision and Pattern Recognition. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1, S. 886–893.
4. N. Zhiheng, S. Shiguang, Y. Shengye, C. Xilin und G. Wen, „2d cascaded

adaboost for eye localization“, in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, Vol. 2, S. 1216–1219.

5. Q. Zhu, M.-C. Yeh, K.-T. Cheng und S. Avidan, „Fast human detection using a cascade of histograms of oriented gradients“, in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, Vol. 2. IEEE, S. 1491–1498.
6. J. Parris, M. Wilber, B. Heflin, H. Rara, A. El-Barkouky, A. Farag, J. Movellan, M. Castrilon-Santana, J. Lorenzo-Navarro und M. N. Teli, „Face and eye detection on hard datasets“, in *Biometrics (IJCB), 2011 International Joint Conference on*. IEEE, S. 1–10.
7. Y. Freund und R. E. Schapire, „A desicion-theoretic generalization of on-line learning and an application to boosting“, in *Computational learning theory*. Springer, 1995, S. 23–37.
8. S. L. Phung und A. Bouzerdoum, „A new image feature for fast detection of people in images“, *Int. J. Inf. Syst. Sci*, Vol. 3, Nr. 3, S. 383–391, 2007.
9. BioID Technology Research, „The bioid face database“, https://www.bioid.com/About/BioID-Face-Database.
10. University of Essex, „Essex face database“, http://cswww.essex.ac.uk/mv/allfaces/.
11. Utrecht University, „Utrecht face database“, http://pics.psych.stir.ac.uk/2D_face_sets.htm.
12. A. S. Georghiades, P. N. Belhumeur und D. Kriegman, „From few to many: Illumination cone models for face recognition under variable lighting and pose“, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 23, Nr. 6, S. 643–660, 2001.
13. L. Fei-Fei, R. Fergus und P. Perona, „One-shot learning of object categories“, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 28, Nr. 4, S. 594–611, 2006.
14. O. Jesorsky, K. J. Kirchberg und R. W. Frischholz, „Robust face detection using the Hausdorff distance“, in *Audio-and video-based biometric person authentication*. Springer, 2001, S. 90–95.

Segmentierung unterschiedlich stark ausgeprägter Welligkeiten auf lackierten Oberflächen

Markus Vogelbacher[1], Mathias Ziebarth[1], Sabine Olawsky[1] und Jürgen Beyerer[1,2]

[1] Karlsruher Institut für Technologie, Institut für Anthropomatik, Lehrstuhl für Interaktive Echtzeitsysteme, Adenauerring 4, D-76131 Karlsruhe
[2] Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung, Fraunhoferstraße 1, D-76131 Karlsruhe

Zusammenfassung Die Wahrnehmung lackierter Oberflächen wird sehr stark durch die vorliegenden Oberflächenrauheiten und -welligkeiten beeinflusst. Entscheidend für den visuellen Eindruck ist dabei nicht nur die Ausprägung der Strukturen an sich, sondern auch in wie weit sich diese über die lackierte Fläche verändern. Lokale Änderungen der Oberflächenqualität können dazu führen, dass solche Oberflächen nicht mehr als homogen wahrgenommen werden. Die Einschätzung der Ausprägung der Welligkeiten wird in der Industrie mit Hilfe verschiedener Messinstrumente durchgeführt. Diese ignorieren dabei meist den zweidimensionalen Charakter des Oberflächeneffektes und tasten das optische Profil der Oberfläche nur linienhaft ab. Änderungen in der Ausprägung werden durch Ausgabe der mittleren Ausprägung bzw. durch Anzeige einer Fehlmessung nicht berücksichtigt. In diesem Beitrag werden Methoden vorgestellt, die zum einen die 2D Aufnahme und Auswertung der Ausprägung der Welligkeiten ermöglichen und zum anderen Änderungen in der Welligkeit detektieren. Dazu wird ein deflektometrisches Messsystem verwendet, um das Gradientenfeld der Oberfläche zu erhalten. Durch Auswertung in bestimmten Wellenlängen-/Frequenzbereichen kann damit sowohl eine Vergleichbarkeit zu Standardkenngrößen aus der Industrie als auch eine Segmentierung unterschiedlich ausgeprägter Welligkeiten erreicht werden.

1 Einleitung

Die Lackschicht einer Oberfläche, wie zum Beispiel aus dem Automobilbereich bekannt, dient neben dem Schutz des darunter liegenden Materials auch zur Beeinflussung der Wahrnehmung der Oberfläche. Dabei ist der visuelle Eindruck der Oberfläche sehr stark durch die auftretenden Oberflächenrauheiten und -welligkeiten bestimmt. Die Welligkeiten bzw. die Überlagerung von Strukturen unterschiedlicher Wellenlängen, die auch als Orangenhaut bezeichnet wird, entsteht durch eine ungleichmäßige Abtrocknung des aufgetragenen Lackes während des Lackierprozesses und liegt in der lateralen Größenordnung zwischen 0,1 mm und 10 mm (Abbildung 12.1). Je nach Ausprägung kann der daraus resultierende Oberflächeneindruck als mehr oder weniger störend beurteilt werden. Außerdem können die auftretenden Welligkeiten zum kaschieren kleinerer Defekte genutzt werden.

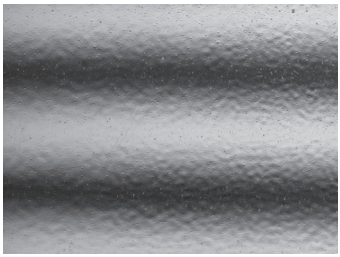


Abbildung 12.1: Visueller Eindruck von Orangenhaut bei Betrachtung eines Streifenmusters.

Insgesamt zeigt sich, dass der Überwachung der Oberflächeneigenschaften eine wichtige Rolle im Bereich der Qualitätskontrolle zuteil wird. Dabei ist es neben der Ableitung von Kenngrößen, die den visuellen Eindruck quantifizieren und dadurch auch Rückschlüsse auf den Lackierprozess ermöglichen sollen, ebenso wichtig die Gleichmäßigkeit der Ausprägung der Welligkeiten über die komplette Oberfläche zu überwachen. Eine lokale Änderung der Ausprägung und damit einhergehend eine Änderung der Abbildungsqualität führt zu einer inhomogenen Wahrnehmung der Oberfläche und kann auch als Defekt angesehen werden. Solche lokalen Änderungen können zum Beispiel durch

ungleichmäßiges Aufbringen des Lackes, beim Polieren der Lackoberfläche oder durch Beschädigungen auftreten.

Im Folgenden wird zunächst in Abschnitt 2 ein Überblick über den Stand der Technik im Bereich der Beurteilung von lackierten Oberflächen gegeben und in diesem Zusammenhang das in der Industrie häufig verwendete Messgerät wave-scan vorgestellt. Abschnitt 3 beinhaltet das verwendete deflektometrische Messprinzip und die jeweiligen Methoden zur Ableitung einer Kenngröße zur Einschätzung der Ausprägung der Welligkeiten und zur Segmentierung unterschiedlich ausgeprägter Welligkeiten. Die Ergebnisse werden in Abschnitt 4 diskutiert.

2 Stand der Technik

Für die Beurteilung spiegelnder Oberflächen bzw. lackierter Bleche existieren verschiedene Kenngrößen, die die Klarheit der Reflexion beschreiben. Dazu wird die Streuung eines Lichtpunktes nach der spiegelnden Reflexion betrachtet [1, 2]. In diesem Zusammenhang häufig verwendete Kenngrößen beschreiben den Glanzverlust (GLOSS, [1]), Kontrast- bzw. Schärfeverlust (HAZE, [1]) und die Verzerrung des reflektierten Bildes (DOI, [2]). Eine Übersicht wie spiegelnde Oberflächen auf Basis von Lichtstreuung und strukturierten Licht charakterisiert werden können liefert Tian et al. [3]. Pietschmann [4] gibt eine allgemeine Übersicht zur Messung der Wahrnehmung bei lackierten Oberflächen. Weitere Arbeiten verwenden unterschiedlichste Aufnahmesysteme als Grundlage für die Orangenhautbewertung [5–8]. Außerdem sind in der ISO 25178-2:2012 3D Oberflächenkenngrößen definiert, die auch zur Beurteilung herangezogen werden können [9].

Verfahren, die speziell eine Beurteilung in definierten Wellenlängenbereichen verwenden, werden häufig in der Industrie eingesetzt und orientieren sich dabei an der menschlichen Wahrnehmung. Ein einfaches, wenn auch sehr grobes Beispiel dafür ist die Unterteilung des Oberflächenprofils in Shortterm Waviness (SW), für alle Strukturen mit Wellenlängen kleiner als 0,6 mm und Longterm Waviness (LW), für Strukturen mit Wellenlängen zwischen 0,6 und 100 mm. Detaillierter fällt die Betrachtung bei dem von Byk-Gardner speziell zur Beurteilung von Orangenhaut entwickelten Messgerät

wave-scan [10] aus. Dieses Gerät tastet durch Messung der über die Oberfläche reflektierten Lichtintensität einer Punktlichtquelle das optische Profil der Oberfläche eindimensional ab und unterteilt mittels mathematischer Filter das Profil in die für die Beurteilung von Orangenhaut relevanten Wellenlängenbereiche von $\lambda_a = 0,1 \ldots 0,3$ mm, $\lambda_b = 0,3 \ldots 1$ mm, $\lambda_c = 1 \ldots 3$ mm, $\lambda_d = 3 \ldots 10$ mm bis $\lambda_e = 10 \ldots 30$ mm ein. Für Wellenlängenbereiche kleiner 0,1 mm (λ_{du}) wird zusätzlich die Lichtstreuung durch betrachten der Reflexion einer LED mit sehr flachem Einfallswinkel in Bezug zur Flächennormalen bestimmt. In einer vorangegangenen Arbeit wurde diese Art der Kenngrößenbestimmung auch für 2D-Gradientenbilder aus einer deflektometrischen Messung abgeleitet [11].

Änderungen in der Ausprägung der verschiedenen Wellenlängen werden bei den bisher bekannten Verfahren nicht betrachtet.

3 2D-Beurteilung lackierter Oberflächen

3.1 Deflektometrie

Das deflektometrische Messprinzip ermöglicht es Gestaltinformationen spiegelnder Oberflächen zu gewinnen [12, 13]. Dabei zeichnet es sich insbesondere durch eine hohe Empfindlichkeit bezüglich lokaler Oberflächenneigungen aus und ist damit sehr gut geeignet, um die hier vorliegenden wellenartigen Lackstrukturen festzuhalten. Das Messsystem besteht aus einer Kamera *I*, der spiegelnden Oberfläche *S* und einem Schirm *L* (Abbildung 12.2). Die Kamera beobachtet über die spiegelnde Oberfläche den Schirm, auf dem eine Sequenz aus phasenverschobenen horizontalen und vertikalen Sinusmustern abgebildet wird. Dadurch kann jedem Kamerapixel eindeutig ein Schirmpixel zugewiesen werden und die entstehenden Sichtstrahlen hängen nur von der Oberflächennormalen ab. Diese Zuordnung nennt sich deflektometrische Registrierung. Auf Basis dieser Information können die 2D-Gradientenfelder $\frac{\partial g(x,y)}{\partial x}$ und $\frac{\partial g(x,y)}{\partial y}$ der Oberfläche $g(x, y)$ erhalten werden.

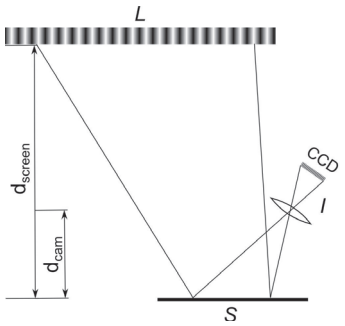


Abbildung 12.2: Deflektometrisches Messsystem bestehend aus Kamera I, spiegelnder Oberfläche S und Schirm mit Sinusmuster L.

3.2 Kenngrößen zur Beschreibung der Oberflächencharakteristik

Aus der Fouriertransformation der aus der deflektometrischen Messung gewonnenen Gradientenfelder $\frac{\partial g(x,y)}{\partial x}$ und $\frac{\partial g(x,y)}{\partial y}$ (Abbildung 12.3) können durch Verwendung des Winkelleistungsspektrums (APS) Kenngrößen zur Beschreibung des Orangenhauteffektes gewonnen werden [11]. Nach ISO 25178-2:2012 ist das Winkelleistungsspektrum für ein Frequenzband $f_i = 1/\lambda_i$ definiert als

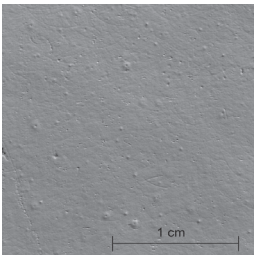
$$\mathrm{APS}_i(\vartheta) = \int_{f_i} r \left|\mathcal{F}\{g\}(r\sin(\vartheta), r\cos(\vartheta))\right|^2 \mathrm{d}r. \tag{12.1}$$

r und ϑ beschreiben die Polarkoordinaten zu den zugehörigen kartesischen Koordinaten f_x und f_y des Frequenzspektrums. Durch Verwendung der Fourierkorrespondenz der Ableitung

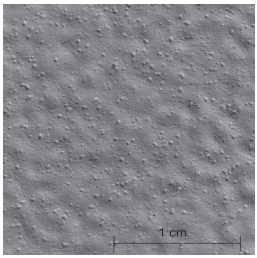
$$G_x(f_x, f_y) := \mathcal{F}\Big\{\frac{\partial g(x,y)}{\partial x}\Big\} = \mathrm{j}2\pi f_x \mathcal{F}\{g(x,y)\}, \tag{12.2}$$

ergibt sich

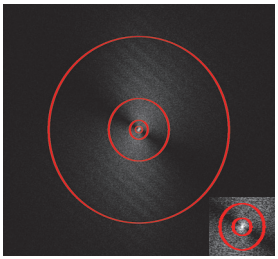
$$\begin{aligned}\mathrm{APS}_i(\vartheta) = \int_{f_i} \frac{1}{4\pi^2 r}\Big(&\left|G_x(r\sin(\vartheta), r\cos(\vartheta))\right|^2 \\ &+ \left|G_y(r\sin(\vartheta), r\cos(\vartheta))\right|^2\Big)\,\mathrm{d}r\end{aligned} \tag{12.3}$$



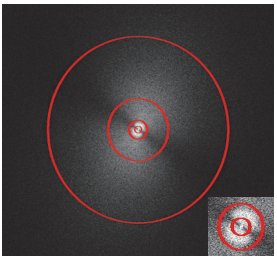
(a) lackierte Oberfläche ohne Orangenhaut



(b) lackierte Oberfläche mit Orangenhaut



(c) Frequenzspektrum für Oberfläche ohne Orangenhaut



(d) Frequenzspektrum für Oberfläche mit Orangenhaut

Abbildung 12.3: Gradientenbilder von lackierten Oberflächen ohne (a) und mit Orangenhaut (b) und zugehörige Frequenzspektren (c,d). Rote Kreise zeigen die Grenzfrequenzen $1/0{,}1\,\mathrm{mm}$, $1/0{,}3\,\mathrm{mm}$, $1/1\,\mathrm{mm}$, $1/3\,\mathrm{mm}$ und in den vergrößerten Ausschnitten rechts unten in (c) und (d) die Grenzfrequenzen $1/1\,\mathrm{mm}$ und $1/3\,\mathrm{mm}$.

Da eine Vorzugsrichtung nicht zu erkennen und auch nicht von Interesse ist bei der Betrachtung von Orangenhaut, kann über alle Winkel ϑ integriert werden

$$\mathrm{BPS}_i = \int_0^{2\pi} \mathrm{APS}_i(\vartheta)\,\mathrm{d}\vartheta. \tag{12.4}$$

Werden diese Kenngrößen für die Frequenzbereiche des wave-scan bestimmt, also $i \in \{du, a, b, c, d, e\}$, so können die erhaltenen BPS Werte auf die Kenngrößen des wave-scan übertragen werden [11] und damit

eine Beurteilung ermöglicht, ob eher lang- oder kurzwellige Strukturen die Oberfläche dominieren.

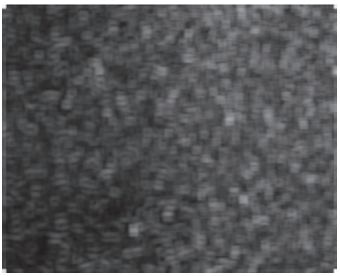
3.3 Segmentierung von Welligkeitsänderungen

Im Bereich der Texturanalyse würde das Gradientenfeld einer Oberfläche mit den betrachteten Welligkeiten dem statistischen Texturtyp zugeordnet werden. Aus diesem Grund kann man zur Detektion von Bereichen unterschiedlich stark ausgeprägter Welligkeiten (Abbildung 12.4(a)) verschiedene statistische Kenngrößen heranziehen. Zuvor ist es allerdings notwendig Fokusunterschiede, die sich aus der Objekt-Kamerakonstellation ergeben, in den Messungen auszugleichen. In Abbildung 12.4(a) zeigt sich dieser Effekt durch einen verrauschten Bereich in der mittleren Horizontalen des Gradientenbildes. Um dies auszugleichen, kann ein Medianfilter angewandt werden (Abbildung 12.4(b)). Für das so bereinigte Gradientenbild wird im nächsten Schritt die lokale Standardabweichung innerhalb einer definierten Nachbarschaft bestimmt (Abbildung 12.4(c)). Kurzwellige Bereiche führen in diesem Zusammenhang zu einer größeren lokalen Standardabweichung. Durch Anwendung eines Mittelwertfilters entsprechender Größe können Bereiche mit ähnlicher Welligkeitsausprägung zusammengefasst werden (Abbildung 12.4(d)). Dadurch können Ausprägungsverläufe sichtbar gemacht werden und je nach Vorgabe unterschiedliche Bereiche definiert werden.

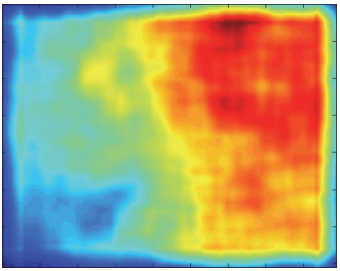
Wie in Abschnitt 2 angedeutet repräsentiert die Betrachtung unterschiedlicher Frequenzbereiche auch in gewisser Weise die menschliche Wahrnehmung. Mit Hilfe der zuvor definierten Frequenzbereiche und der Wavelettransformation kann dies auch für die Segmentierung eingebracht werden. Nach Unterdrückung kleinerer Defekte, wird auf das Gradientenfeld eine kontinuierliche Wavelettransformation durchgeführt [14]. Als Wavelet wird in diesem Fall ein Mexican Hat Wavelet verwendet und über die Pseudofrequenz der Waveletfunktion die betrachteten Skalierungen an bestimmte Frequenzen angepasst. Konkret werden die Frequenzen enstprechend den Grenzen bzw. der Mitte der Frequenzbereiche des wave-scan gewählt, nämlich $10\,{}^{1}/_{\mathrm{mm}}, 6,67\,{}^{1}/_{\mathrm{mm}}, 2,17\,{}^{1}/_{\mathrm{mm}}, 0,67\,{}^{1}/_{\mathrm{mm}}, 0,22\,{}^{1}/_{\mathrm{mm}}$ und $0,1\,{}^{1}/_{\mathrm{mm}}$. Anschließend wird auf den erhaltenen Skalen die lokale Energie in einer der Skalierung angepassten Nachbarschaft berechnet. Nach

(a) Gradientenfeld der betrachteten Oberfläche

(b) Anwendung eines Medianfilters zur Beseitigung von Fokusunterschieden



(c) Bestimmung der lokalen Standardabweichung



(d) Detektionsergebnis nach Mittelwertfilterung

Abbildung 12.4: Beispiel für die Detektion von Bereichen unterschiedlich stark ausgeprägter Welligkeiten mittels lokaler Standardabweichung. (d) zeigt in einer Falschfarbendarstellung den Übergang von links, dem langwelligen Bereich (blau), nach rechts, dem kurzwelligen Bereich (rot) der betrachteten Oberfläche.

Glättung der Energiebilder durch einen Gaußfilter, können letztendlich Bereiche mittels eines Region Growing Algorithmus [15] segmentiert werden (Abbildung 12.5).

4 Diskussion der Ergebnisse

In [11] kann die Eignung der gewonnenen 2D-Kenngrößen aus der deflektometrischen Messung nachvollzogen werden. Es wird ein Vergleich zwischen den BPS Werten und den Kenngrößen der wave-scan Mes-

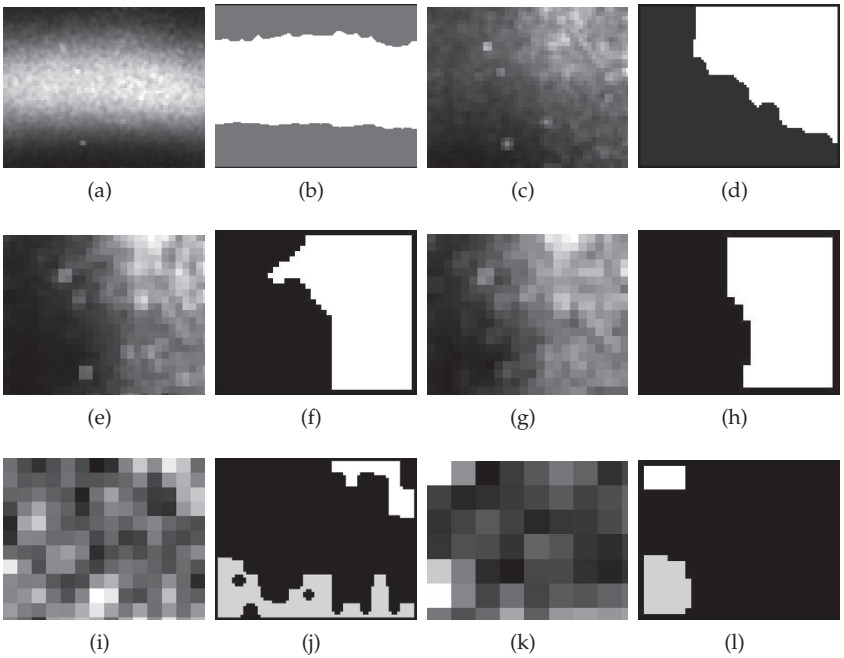
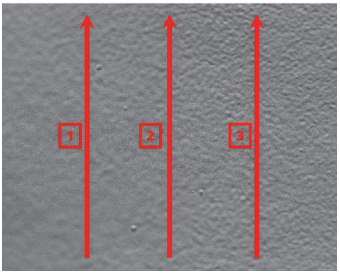


Abbildung 12.5: Beispiel für die Detektion von Bereichen unterschiedlich stark ausgeprägter Welligkeiten mittels kontinuierlicher Wavelettransformation für das Gradientenfeld aus Abbildung 12.4(a) und den entsprechenden Pseudofrequenzen der Skalen (a,b) $10\,{}^{1}/_{\mathrm{mm}}$, (c,d) $6,67\,{}^{1}/_{\mathrm{mm}}$, (e,f) $2,17\,{}^{1}/_{\mathrm{mm}}$, (g,h) $0,67\,{}^{1}/_{\mathrm{mm}}$, (i,j) $0,22\,{}^{1}/_{\mathrm{mm}}$ und (k,l) $0,1\,{}^{1}/_{\mathrm{mm}}$. Für die jeweiligen Skalen ist links die lokale Energie und rechts das Segmentierungsergebnis dargestellt.

sung für Testbleche mit unterschiedlichen Glanzgraden und Orangenhautausprägungen durchgeführt.

Auch für die Segmentierung können wave-scan Messungen herangezogen werden, um die Korrektheit der Ergebnisse zu bestätigen. Dazu sind allerdings mehrere solcher wave-scan Messungen notwendig. Wird für die Oberfläche aus Abbildung 12.4(a) wie in Abbildung 12.6(a) dargestellt an drei Stellen eine wave-scan Messung durchgeführt, so sind deutliche Veränderungen in den vom wave-scan gelieferten Kenn-



(a) Referenzierung der wave-scan Messungen auf die deflektometrische Messung

	B	W_{du}	W_a	W_b	W_c	W_d	W_e
1	6,9	60,5	68	61,3	32	29,1	19,1
2	6,5	61,3	69,7	65	46,6	35	18
3	8,3	62,2	69,7	72,3	56,4	35	25,3

(b) Kenngrößen der wave-scan Messungen

Abbildung 12.6: Auswertung des Oberflächenabschnittes aus Abbildung 12.4(a) mittels wave-scan. W_{du} bis W_e beschreiben die Kenngrößen der entsprechenden Wellenlängenbereiche. B wird als Balancewert bezeichnet, ein von Byk-Gardner aus einer Studie hergeleitete Kenngröße, die die Kenngrößen W_b und W_d kombiniert, um die subjektive Wahrnehmung der Orangenhaut zu beschreiben.

größen für diese Messbereiche und damit in den unterschiedlichen Wellenlängebereichen zu erkennen (Abbildung 12.6(b)). Möchte man eine feinere Auflösung der Änderung in der Ausprägung, wären mehr Messpunkte notwendig. Durch Verwendung der 2D deflektometrischen Messung und mit Hilfe der lokalen Standardabweichung können diese Ergebnisse in einer Messung gewonnen werden (Abbildung 12.4(d)). Für eine automatische Auswertung können damit Bereiche definiert werden, in denen eine Kenngrößenschwankung nur in einem vorher festgelegten Bereich vorliegen. Somit ist eine sehr feine Auflösung möglich und es wird verhindert, dass Ausprägungsänderungen herausgemittelt werden, wie es zum Beispiel beim wave-scan der Fall wäre. Bei der Betrachtung der Segmentierung auf unterschiedlichen Skalen der Wavelettransformation zeigt sich, dass vor allem die Pseudofrequenzen bei $2,17\,{}^{1}/_{\mathrm{mm}}$ und $0,67\,{}^{1}/_{\mathrm{mm}}$ (Abbildung 12.5 (e-h)) für eine Segmentierung unterschiedlicher Wellenlängenausprägungen geeignet sind. Die Skala der Pseudofrequenz bei $10\,{}^{1}/_{\mathrm{mm}}$ (Abbildung 12.5 (a,b)) könnte wiederum verwendet werden, um den Kamerafokus zu bestimmen. Die übrigen Skalen zeigen sich als ungeeignet für die Bestimmung von Ausprägungsänderungen.

5 Zusammenfassung

In diesem Beitrag werden Verfahren vorgestellt, die es ermöglichen verschiedene Ausprägungen der Orangenhaut auf lackierten Blechen festzustellen und zu segmentieren. Da dies ein zweidimensionaler Effekt ist, wird im Gegensatz zu den bekannten Messinstrumenten ein flächiges Messverfahren, die Deflektometrie, verwendet. Auf Grundlage der aus der Deflektometrie erhaltenen Gradientenbilder der Oberfläche können zum einen über das Frequenzspektrum und das Winkelleistungsspetrum die 1D Kenngrößen aus der Industrie bekannter Orangenhautmessgeräte abgeleitet und zum anderen Ausprägungsänderungen der vorhandenen Wellenlängen detektiert werden. Für die einfache Detektion von Ausprägungsänderungen ist es dabei ausreichend die lokale Standardabweichung des Gradientenbildes auszuwerten. Eine wellenlängenabhängige Segmentierung kann mittels der Betrachtung der Skalen einer kontinuierlichen Wavelettransformation für bestimmte Pseudofrequenzen und der Auswertung der lokalen Energie erreicht werden.

Aus der Kombination von Kenngrößenbestimmung und Detektion von Bereichen unterschiedlich ausgeprägter Wellenlängen wird es damit möglich in einer Messung flexibel Bereiche gleicher Ausprägung festzulegen und die entsprechenden Kenngrößen anzugeben. Lokale Abweichungen von einer ansonsten homogenen Lackschicht können somit als Fehler detektiert werden. Insgesamt wird dadurch die Auswertung von Kenngrößen auf Grund der 2D-Daten schneller und robuster. Eine gleichwertige Auswertung der Oberfläche zum Beispiel mit dem wave-scan würde eine Vielzahl an Messungen benötigen und könnte auf Grund der Mittelwertbildung über einen definierten linienhaften Oberflächenabschnitt zu Fehlern führen.

Literatur

1. G. Kigle-Boeckler, „Measurement of gloss and reflection properties of surfaces“, *Metal Finishing*, S. 28–31, 1995.
2. M.-K. Tse, D. Forrest und E. Hong, „An improved method for distinctness of image (doi) measurements“, Quality Engineering Associates (QEA), Inc., Burlington, MA, USA, Tech. Rep., 2005.

3. G. Tian, R. Lu und D. Gledhill, „Surface measurement using active vision and light scattering“, *Optics and Lasers in Engineering*, Vol. 45, Nr. 1, S. 131–139, 2007.
4. J. Pietschmann, *Industrielle Pulverbeschichtung*. Vieweg+Teubner, 2010, Vol. 3.
5. T. Fletcher, „A simple model to describe relationships between gloss behaviour, matting agent concentration and the rheology of matted paints and coatings“, *Progress in Organic Coatings*, Vol. 44, S. 25–36, 2002.
6. M. Miranda-Medina, T. Wagner, J. Böhm, A. Vernes und K. Hingerl, „Optical analysis of orange peel on metallic surfaces“, in *Proceedings of SPIE Optical Micro- and Nanometrology IV*, 2012.
7. M. Osterhold, „Characterization of surface structures by mechanical and optical fourier spectra“, *Progress in Organic Coatings*, Vol. 27, S. 195–200, 1996.
8. A. F. R.-S. Lu, „3d surface topography from the specular lobe of scattered light“, *Optics and Lasers in Engineering*, Vol. 45, S. 1018–1027, 2007.
9. S. Rebeggiani, B.-G. Rosén und A. Sandberg, „A quantitative method to estimate high gloss polished tool steel surfaces“, *Journal of Physics: Conference Series*, Vol. 311, Nr. 1, 2011.
10. „BYK-Gardner GmbH, Orange Peel and DOI Meters“, https://www.byk.com/en/instruments/products/appearance-measurement/orange-peel-doi-meter.html, accessed on 2014-02-13.
11. M. Ziebarth, M. Vogelbacher, S. Olawsky und J. Beyerer, „Obtaining 2d surface characteristics from specular surfaces“, in *German Conference on Pattern Recognition*, 2014.
12. D. Pérard, „Automated visual inspection of specular surfaces with structured-lightning reflection techniques“, Dissertation, University of Karlsruhe, 2000.
13. S. Werling, M. Mai, M. Heizmann und J. Beyerer, „Inspection of specular and partially specular surfaces“, *Metrology and Measurement Systems*, Vol. 16, S. 415–431, 2009.
14. S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, 3. Aufl. Academic Press, 2008.
15. R. Adams und L. Bischof, „Seeded region growing“, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 16, Nr. 6, S. 641–647, Jun 1994.

Digitalisierung dreidimensionaler CFK-Halbzeuge zur Fehlstellenklassifizierung am Beispiel der Faserwelligkeit

Philipp Kosse, Tobias Fürtjes und Robert Schmitt

RWTH Aachen,
Lehrstuhl für Fertigungsmesstechnik und Qualitätsmanagement am Werkzeugmaschinenlabor WZL,
Steinbachstraße 19, 52074 Aachen

Zusammenfassung Durch unvermeidbare Schwankungen im Drapierprozess dreidimensionaler kohlenstofffaserverstärkter Kunststoff-Halbzeuge (CFK-Preforms) kann eine optische Prüfung auf textile Fehlstellen nur erfolgen, wenn neben der Oberfläche auch die unbekannte Geometrie erfasst wird. Diese Arbeit zeigt daher eine Methode zur Digitalisierung dreidimensionaler Halbzeuge, die die schwierigen optischen Eigenschaften des Materials überwindet. Dabei werden die Datenpunkte eines Laser-Lichtschnittsensors mit der optisch detektierten Faserorientierung in einem gemeinsamen Datenmodell fusioniert. Mithilfe des Datenmodells erfolgt anschließend die Fehlstellenerkennung und Fehlstellenklassifizierung am Beispiel der Faserwelligkeit in der Bauteilebene und in Bauteildicke.

1 Einleitung

Kohlenstofffaserverstärkte Kunststoffe (CFK) setzen sich im industriellen Einsatz als Leichtbaualternative gegenüber konventionellen metallischen Werkstoffen immer weiter durch. Gerade im Bereich Automotive, Aerospace und Windenergie wird der Werkstoff immer häufiger eingesetzt.

Bauteile aus CFK werden aus zwei Komponenten gefertigt: Carbonfasern und ein Matrix-Werkstoff, der die Fasern in Position hält. Dabei wird die Festigkeit und Steifigkeit der Bauteile maßgeblich durch

die Orientierung der Fasern bestimmt, entlang derer auf das Bauteil wirkende Kräfte aufgenommen werden können. Diese Anisotropie ist für die Fertigung dreidimensionaler Bauteile eine besondere Herausforderung. Trotz einer Vielzahl unterschiedlicher Fertigungsverfahren ist bei keinem Fertigungsprozess für dreidimensionale Bauteile eine hinreichend hohe Prozessstabilität hinsichtlich der Faserorientierung in naher Zukunft zu erwarten. Ein Grund hierfür ist die fehlende Qualitätssicherung in den Fertigungsprozessen.

Diese Arbeit beschreibt eine Methode zur vollständigen optischen Prüfung textiler Halbzeuge mithilfe eines 3D-Sensorsystems, geführt von einem Industrieroboter. Das beschriebene System ist dabei nicht an ein spezielles Fertigungsverfahren gekoppelt, sondern kann für alle Arten textiler Halbzeuge eingesetzt werden. Die Herausforderungen bestehen hierbei in der Überwindung der schwierigen optischen Eigenschaften der schwarz-glänzenden Carbonfasern und in einer automatisierten Fehlererkennung und Fehlerklassifizierung, die am Beispiel der Faserwelligkeit evaluiert werden.

2 Stand der Technik

Drapieren beschreibt den Prozess, ein zweidimensionales flächiges Textil in eine dreidimensionale Oberfläche umzuformen [1]. Die Qualität des dabei entstehenden Faserhalbzeuges (Preform) ist wegen der anisotropen Eigenschaften von CFK-Bauteilen nicht nur durch die lokal eingehaltene Faserorientierung bestimmt, sondern auch durch Vermeidung von Drapierfehlern, wie z. B. Gassen, Falten und der Faserwelligkeiten [2]. Selbst bei Umformungsgraden, die nicht über den maximalen Grenzscherwinkel hinaus gehen, ist eine Fehlstellenbildung nicht auszuschließen [3]. Ohne das Leichtbaupotenzial durch eine Überdimensionierung zu verlieren, kann der mangelnden Prozessstabilität nur mit Methoden der Qualitätssicherung entgegengewirkt werden.

Für die Qualitätssicherung von CFK eignen sich ausschließlich zerstörungsfreie Prüfmethoden. Diese lassen sich in zwei Kategorien einordnen: zum einen Prüfsysteme für konsolidierte Bauteile; zum anderen für textile Halbzeuge. Gängige Verfahren der ersten Kategorie sind u. a. Ultraschall, Thermografie, Wirbelstromprüfung, Shearografie und die Computertomografie [4,5]. Für textile Halbzeuge sind dagegen

überwiegend optische Systeme relevant [6], ebenso wie die Computertomografie [7].

Obwohl die meisten Systeme der ersten Kategorie prinzipiell zur Erkennung der Faserorientierung geeignet sind, ist der Einsatz am textilen Halbzeug nicht zweckmäßig. CFK-Preforms bestehen aus aufeinandergelegten, (bebinderten) Faserlagen, zwischen denen eine Luftbrücke liegt. Für das Ultraschall- und Thermografie-Verfahren wird die Signalausbreitung durch die Luft zwischen den Lagen dämpft und ist somit nicht mehr wirkungsvoll am Preform einsetzbar. Eine Wirbelstromprüfung ist aufgrund der begrenzen Eindringtiefe ebenfalls durch die Luftbrücken eingeschränkt. Zusätzlich nachteilig ist die punktweise Prüfung analog zum klassischen Ultraschall. Die flexible, biegeschlaffe Struktur verhindert den Einsatz der Shearografie, die zudem nicht die Faserorientierung erfassen kann.

Am besten geeignet für die Prüfung von CFK-Preforms sind optische Sensoren [8]. Verschiedene Varianten der optischen Prüfung verwenden eine spezielle Dom- oder Flächenbeleuchtung in Kombination mit einer hochauflösenden Industriekamera. Die Beleuchtung kann entweder diffus für eine homogene Ausleuchtung ausfallen [6, 9] oder sequenziell gerichtet, bei der nacheinander verschiedene Sektoren einer Dombeleuchtung einzeln geschaltet werden [10]. Zur Ermittlung der Faserorientierung hat sich das Strukturtensorverfahren etabliert [6].

Zur Erfassung der Geometrie von Carbonfasern haben sich laserbasierte Lichtschnittverfahren als zuverlässig erwiesen. Eine Laserlinie wird dabei auf das Preform projiziert, die von einer Kamera unter einem definierten Winkel aufgenommen wird. Mithilfe der Photogrammetrie wird aus dem beobachteten Versatz der Laserlinie das darunterliegende Höhenprofil errechnet. In [11] wurde die Methode verwendet, um die Kontur von CFK-Preforms zu messen. In [12] wird das Laser-Lichtschnittverfahren zur Erkennung der Anzahl abgelegter Faserlagen und zur Erkennung von Falten eingesetzt. Streifenprojektionssysteme, die bereits industriell für Objekte aus konventionellen Materialien eingesetzt werden, sind bisher nicht in der Lage Carbonfaseroberflächen zu scannen.

Ein umfassendes Konzept zur Qualitätssicherung von CFK-Preform wird in [13] vorgestellt. Ein Laser-Lichtschnittsensor und eine Kamera sollen dabei von einer dreiachsigen Portalanlage über das zu scannende Preform geführt werden. Realisiert wurde das System bisher nicht.

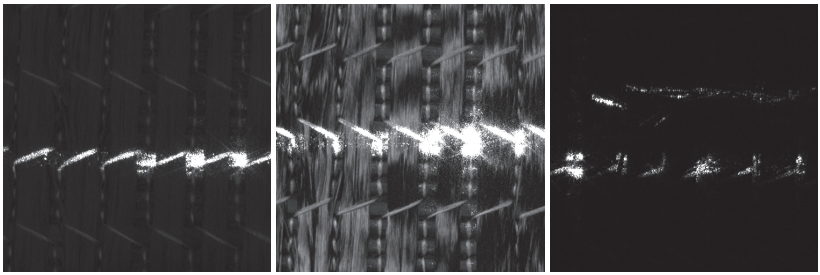
Abbildung 13.1: Robotergeführtes Kamerasystem mit Laser-Lichtschnittsensor.

Ein anderes, robotergeführtes Konzept wird in [9] vorgestellt, bestehend aus einem Kamerasystem mit diffuser Beleuchtung und einem Laser-Abstandssensor. Die ermittelten Daten sollen über ein vorhandenes CAD-Modell gelegt werden, bzw. der Abstand zwischen Roboter und Preform soll ausgewertet werden. Begleitend zu dem vorgeschlagenen Konzept werden die Vorteile einer frühen Fehlererkennung angesprochen [9]. Eine Digitalisierung bei unbekannten Geometrien wird nicht umgesetzt.

3 Digitalisierung und Datenfusion

Für die Digitalisierung wurde in dieser Arbeit ein Sensor entwickelt, der aus einer 5-Megapixel-Kamera, einer diffusen Flächenbeleuchtung und einem Linienlaser besteht. Alle Komponenten werden von einem Industrieroboter geführt, wie in Abbildung 13.1 dargestellt. Die maximale Scangröße ist nur durch den Arbeitsbereich des verwendeten Roboters begrenzt.

Die Digitalisierung der CFK-Preforms mithilfe des Roboters erfolgt in zwei Schritten. In einer ersten Fahrt über das Bauteil wird mit der Laser-Linienprojektion die Geometrie des Preforms erfasst. Bei der zweiten Fahrt wird die diffuse Beleuchtung in Kombination mit der Kamera zur Erfassung der Oberfläche verwendet. Die Bahn des Roboters ist entsprechend der Oberfläche des Preforms einprogrammiert, so dass sich die Preformoberfläche stets im Tiefenschärfebereich der Kamera befindet.



(a) Normalfall (b) Lange Belichtungszeit (c) Faserflug

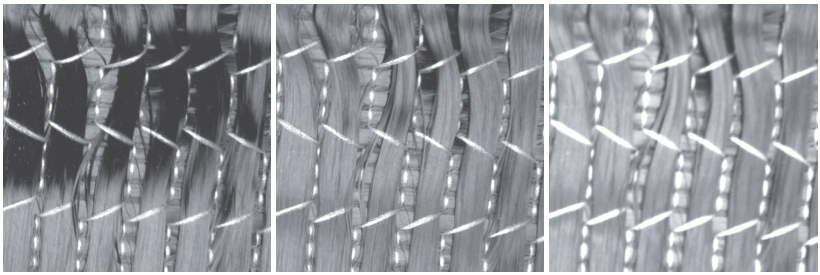
Abbildung 13.2: Ausschnitte der beobachteten Laserlichtschnitte.

3.1 Geometrieerfassung

Abbildung 13.2(a) zeigt den vergrößerten Ausschnitt einer erfassten Laserlinie unter Normalbedingungen. Auffällig ist, dass das Laserlicht fast ausschließlich von den weißen Nähfäden reflektiert wird, kaum aber von den schwarzen Carbonfasern. Eine Erhöhung der Belichtungszeit, wie in Abbildung 13.2(b) gezeigt, ist nicht geeignet die Sichtbarkeit der Laserlinien auf den Carbonfasern zu erhöhen. Stattdessen erschwert der zusätzlich sichtbare Hintergrund die Auswertung der Laserlinie und erfordert eine Segmentierung zur Filterung des Hintergrunds. Desweiteren erschwert die Überbelichtung an den Nähfäden die Auswertung. Abbildung 13.2(c) zeigt das resultierende Bild einer Laserlinie am Rande des Preforms, unter der einzelne Fasern durch den Preformbeschnitt herausragen. In Abhängigkeit der Ausrichtung zum Laser und der Kamera kann eine Reflexion der Laserlinie entstehen. Da die Laserlinie wesentlich breiter im Vergleich zu dem Radius der Carbonfasern ist, entsteht eine zweite parallel verlaufende Reflexion. Für alle weiteren Untersuchungen in Kapitel 4 wurden daher abstehende Fasern entfernt.

3.2 Texturerfassung

Die Erfassung der Textur erfordert trotz der diffusen Beleuchtung eine höhere Belichtungszeit als bei der Geometrieerfassung. Um eine Bewegungsunschärfe in den Bildaufnahmen auszuschließen, hält der Roboter für jede Bildaufnahme vollständig an. Der Bereich einer Einzelauf-



(a) Kameraposition 1 (b) Kameraposition 2 (c) Kameraposition 3

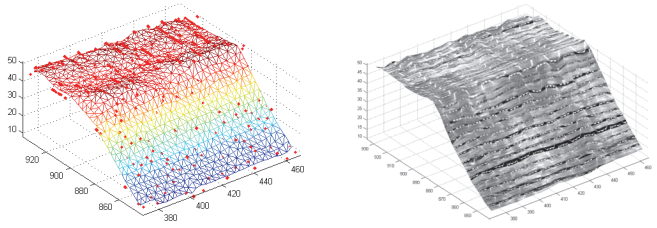
Abbildung 13.3: Bildausschnitte mit positionsabhängiger Abschattung und Unschärfe.

nahme beträgt ca. 70 mm $\times$ 50 mm. Trotz einer diffusen Beleuchtung, die mit 150 mm $\times$ 150 mm wesentlich größer ist als der Bildbereich, entstehen je nach Position des Sensors zur Oberfläche unterschiedliche Abschattungen. Zusätzlich können durch die vorab nicht bekannten Geometrie-Fehler, wie z. B. Falten, weitere Bildbereiche außerhalb des Tiefenschärfebereichs liegen, wie in Abbildung 13.3(c) dargestellt ist. Durch eine hinreichend enge Überlappung der Einzelaufnahmen können diese Nachteile jedoch bei der Datenfusion überwunden werden.

3.3 Datenfusion

Neben den Einzelbildern wird zusätzlich pro Bild auch die aktuelle Roboterposition für jedes Einzelbild aufgezeichnet. Mithilfe der vorab durchgeführten Hand-Auge-Kalibrierung nach [14] können alle Einzelaufnahmen im gemeinsamen Roboter-Basiskoordinatensystem registriert werden. Abbildung 13.4(a) zeigt einen Ausschnitt aus mehreren ermittelten Höhenprofilen und der linear approximierten regelmäßigen Gitterstruktur. Es entsteht eine gerasterte Oberfläche, dessen Maschengröße frei konfigurierbar ist. In der Praxis hat sich eine Maschengröße von 1 mm $\times$ 1 mm als zweckmäßig erwiesen.

Durch die räumliche Überlappung der einzelnen Oberflächenaufnahmen stehen für jede Masche mehrere Texturen zur Verfügung. Basie-



(a) Geometrie aus Lichtschnitten (b) Überlagerte Texturen

Abbildung 13.4: Datenfusion aller Einzelmessungen.

rend auf einem Schwellwertverfahren für die Grauwertverteilung, der Kohärenz der ermittelten Faserorientierungen und des Abstandes der Kamera von der Masche wird eine entsprechende Textur mit möglichst minimaler Abschattung und Unschärfe automatisch ausgewählt. Es entsteht ein fusioniertes Datenmodell des Preforms, wie Abbildung 13.4(b) zeigt. Neben der Oberflächentextur werden auch die ermittelten Orientierungen pro Pixelelement im Datenmodell registriert. Die 3D-Faserorientierung ist somit für jeden Punkt der Oberfläche für eine Fehlererkennung und Fehlerklassifizierung verfügbar.

4 Experimentelle Ergebnisse und Fehlerklassifizierung

Ein Beispiel für die Digitalisierung eines gesamten Preforms ist in Abbildung 13.5 gezeigt. Das Preform mit einer Größe von 500 mm $\times$ 280 mm und einer maximalen Höhe von 70 mm stammt aus Drapierversuchen, eignet sich aber aufgrund der hohen Umformgrade zur Evaluation der Digitalisierung und der anschließenden Fehlerklassifizierung. Die Geometrie wurde aus 1.120 Laser-Lichtschnitten und mit einer Rastergröße von 1 mm $\times$ 1 mm fusioniert. die Oberfläche wurde mit 32 Einzelaufnahmen erfasst. Trotz einer Überlappung von ca. 50 % der Oberfläche ist in diesem Beispiel noch eine deutliche Abschattung einzelner Regionen zu erkennen, insbesondere im Bereich der Falten am Preformrand. In den nicht vollständig schwarzen Bereichen ist dennoch die

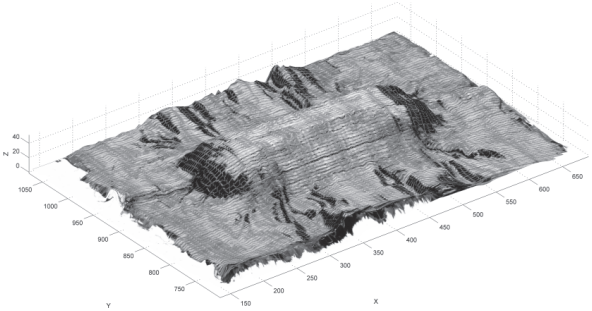


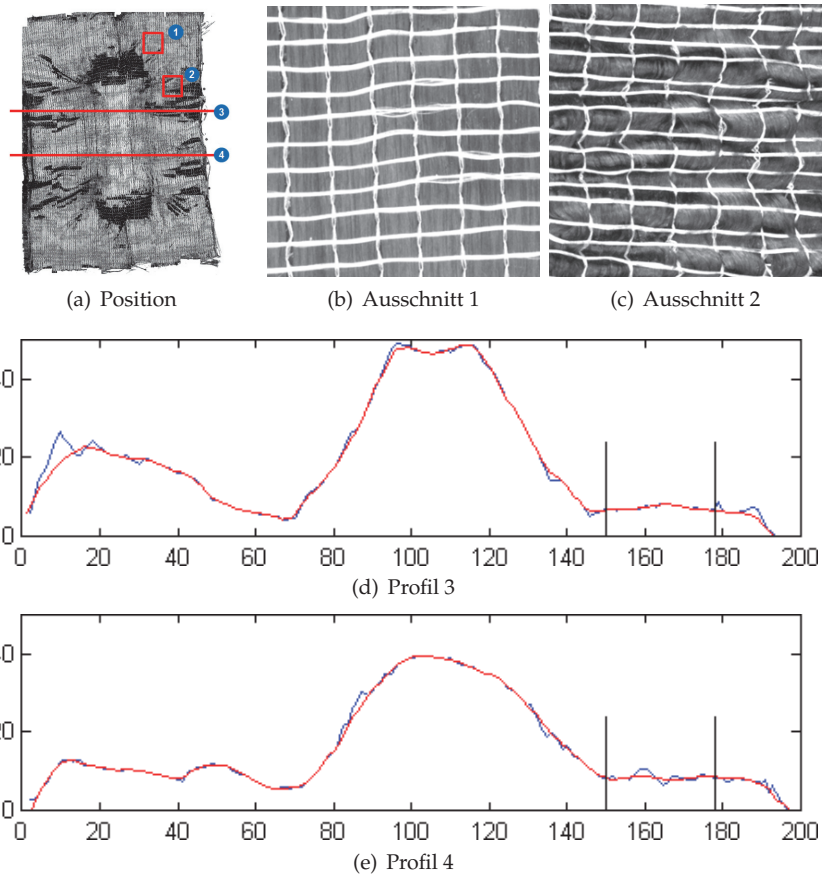
Abbildung 13.5: Digitaliserter Preform

Ermittlung der Faserorientierung über das Strukturtensorverfahren [6] möglich.

Für die Untersuchung der Faserwelligkeit wird das Preform mit einer Rasterung von 30 mm $\times$ 30 mm analysiert. Dabei wird zwischen zwei Fehlerklassen unterschieden: Faserwelligkeit in der Bauteilebene und Faserwelligkeit in Bauteildicke. Für beide Klassen werden exemplarisch jeweils zwei ausgewählte Bereiche des Textils betrachtet, dessen Positionen in Abbildung 13.6(a) dargestellt werden.

Die ersten beiden Ausschnitte, Abbildung 13.6(b) und Abbildung 13.6(c), zeigen die Auswertung der Faserwelligkeit in der Bauteilebene. Innerhalb eines Rasterfeldes werden dazu alle Faserorientierungen der einzelnen Pixelelemente hinsichtlich der Standardabweichung ausgewertet. Für den ersten Fall ohne eine sichtbare Welligkeit beträgt die Standardabweichung $\sigma_1 = 1{,}31°$ bei einem Mittelwert von $\mu_1 = 89{,}21°$; im zweiten Fall mit einer deutlich sichtbaren Welligkeit beträgt die Standardabweichung $\sigma_2 = 43{,}66°$ mit $\mu_2 = 91{,}04°$. Der Mittelwert liefert zwar keinen direkte Hinweis auf eine Welligkeit, kann jedoch für die Erkennung weiterer Fehlerklassen, wie z. B. Falten, genutzt werden.

Die Bereiche 3 und 4 in Abbildung 13.6(d) und Abbildung 13.6(e) zeigen die Höhenprofile entlang der dargestellten Schnitte. Der Soll-Verlauf (rot) der Geometrie wird dabei aus dem detektierten Ist-Verlauf (blau) durch eine starke Glättung der Daten [15] geschätzt. Einen Indikator für die Welligkeit in Bauteildicke liefert die Wurzel des mittleren quadratischen Fehlers (RMSE) zwischen dem Ist-Verlauf und dem ermittelten Soll-Verlauf. Dieser wird für jedes Rasterfeld sowohl in ho-



(a) Position (b) Ausschnitt 1 (c) Ausschnitt 2

(d) Profil 3

(e) Profil 4

Abbildung 13.6: Auswertung der Faserwelligkeit in der Bauteilebene: 13.6(b) bis 13.6(c); Auswertung der Faserwelligkeit in Bauteildicke: 13.6(d) bis 13.6(e).

rizontaler als auch in vertikaler Richtung ausgewertet. Für die in den Abbildungen markierten Bereiche zeichnet sich die sichtbare Welligkeit in dem erhöhten $RMSE_4 = 4{,}18\,\mathrm{mm}$ ab. In Bereichen ohne Welligkeit, wie exemplarisch in Abbildung 13.6(d) dargestellt, bleibt der Fehler geringer, hier bei $RMSE_3 = 1{,}19\,\mathrm{mm}$.

5 Zusammenfassung

In dieser Arbeit wurde ein neues System zur Digitalisierung dreidimensionaler Preforms aus faserverstärkten Kunststoffen vorgestellt, das aus einem robotergeführten optischen Kamerasystem und einem Laser-Lichtschnitt-Sensor besteht. Mit diesem System kann sowohl die Geometrie der Preforms, als auch die Oberfläche erfasst werden, um auf Basis der ermittelten Daten textile Fehlstellen zu detektieren.

Die Fehlerklassifizierung wurde am Beispiel der Faserwelligkeit evaluiert. Dabei kann die Welligkeit in der Bauteilebene durch die Standardabweichung in der Faserorientierung quantifiziert werden. Die Welligkeit in Bauteildicke kann dagegen durch den quadratischen Fehler in der Ist-Geometrie im Vergleich zur geschätzten Soll-Geometrie identifiziert werden. Die Klassifizierung erfolgt durch durch Festlegung von geeigneten Schwellwerten, wobei das Preform in gerasterter Form ausgewertet wird. Die Ermittlung der Schwellwerte ist material- bzw. anwendungsspezifisch und kann nur durch externe Simulationen oder Belastungstest erfolgen.

Noch unbekannt ist die erreichte Messunsicherheit in der Geometrie und der 3D-Faserorientierung des vorgestelten Systems. Forschungsbedarf besteht dabei insbesondere in der Fertigung geeigneter CFK-Prüfkörper.

Insgesamt wurde gezeigt, dass das vorgestellte System eine automatisierte Erkennung und Klassifizierung textiler Fehlstellen ermöglicht. Die Klassifizierung beschränkt sich nicht auf das Beispiel der Faserwelligkeit, sondern kann auf weitere Fehlerklassen erweitert werden.

Literatur

1. M. Flemming, G. Ziegmann und S. Roth, *Faserverbundbauweisen. Halbzeuge und Bauweisen*. Springer, 1996.
2. S. Sharma, M. Sutcliffe und S. Chang, „Characterisation of material properties for draping of dry woven composite material“, *Composites Part A: applied science and manufacturing*, Vol. 34, Nr. 12, S. 1167–1175, 2003.
3. V. Eckers und T. Gries, „Entwicklung eines Prüfplans für Bewehrungen für Textilbeton“, in *Bautechnik*, Vol. 89, Nr. 11. Wiley Online Library, 2012, S. 754–763.

4. P. Vaara, J. Leinonen *et al.*, „Technology Survey on NDT of Carbon-fiber Composites." Kemi-Tornion ammattikorkeakoulu, 2012.
5. B. Ray, S. Hasan und D. Clegg, „Evaluation of defects in frp composites by ndt techniques", *Journal of Reinforced Plastics and Composites*, 2007.
6. C. Mersmann, *Industrialisierende Machine-Vision-Integration im Faserverbundleichtbau*. Aachen: Apprimus Wissenschaftsverlag, 2012.
7. R. Stoessel, T. Guenther, T. Dierig, K. Schladitz, M. Godehardt, P.-M. Kessling und T. Fuchs, „mu-computed tomography for micro-structure characterization of carbon fiber reinforced plastic (cfrp)", in *American Institute of Physics Conference Series*, Vol. 1335, 2011, S. 461–468.
8. A. Orth, *Entwicklung eines Bildverarbeitungssystems zur automatisierten Herstellung faserverstärkter Kunststoffstrukturen*. Shaker Verlag, 2008.
9. S. Gubertanis, J.-M. Balvers und C. Weimer, „Concept development for inline process control of the preform-lcm production chain", *NDT in Aerospace 2012 - We.2.B.4*, 2012.
10. W. Palfinger, S. Thumfart und C. Eitzinger, „Photometric stereo on carbon fiber surfaces", in *35th Workshop of the Austrian Association for Pattern Recognition, Graz*, 2011.
11. R. Schmitt, A. Orth und C. Niggemann, „A method for edge detection of textile preforms using a light-section sensor for the automated manufacturing of fibre-reinforced plastics", in *Optical Metrology*. International Society for Optics and Photonics, 2007.
12. N. Miesen, R. M. Groves, J. Sinke und R. Benedictus, „Laser displacement sensor to monitor the layup process of composite laminate production", in *SPIE Smart Structures and Materials + Nondestructive Evaluation and Health Monitoring*. International Society for Optics and Photonics, 2013.
13. G. Lanza und D. Brabandt, „Design of a measurement machine for quality assurance of preforms in the CFRP process chain." in *ISMTII 2013*. Aachen: Apprimus-Verl., 2013, S. 255–256.
14. R. Y. Tsai und R. K. Lenz, „A new technique for fully autonomous and efficient 3D robotics hand/eye calibration", *Robotics and Automation, IEEE Transactions on*, Vol. 5, Nr. 3, S. 345–358, 1989.
15. W. S. Cleveland, „Robust locally weighted regression and smoothing scatterplots", *Journal of the American statistical association*, Vol. 74, Nr. 368, S. 829–836, 1979.

Sichtbarkeit von Dellen und Beulen auf spiegelnden Oberflächen

Mathias Ziebarth[1], Michael Heizmann[2,3] und Jürgen Beyerer[1,3]

[1] Karlsruher Institut für Technologie, Institut für Anthropomatik und Robotik, Lehrstuhl für Interaktive Echtzeitsysteme, Adenauerring 4, D-76131 Karlsruhe
[2] Hochschule Karlsruhe – Technik und Wirtschaft, Moltkestraße 30, D-76133 Karlsruhe
[3] Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung, Fraunhoferstraße 1, D-76131 Karlsruhe

Zusammenfassung Während sich Fehler auf matten Oberflächen anhand ihrer Ausdehnungen bewerten lassen, ist bei spiegelnden Oberflächen die Auswirkung eines Defekts auf die spiegelnde Abbildung maßgebend. In diesem Beitrag wird ein Zusammenhang zwischen der Sichtbarkeit eines Defekts für den Menschen und dessen Tiefe, Ausdehnung und Krümmung sowie der Reflektanz der Oberfläche hergestellt. Dazu wird eine Studie vorgestellt, bei der die Studienteilnehmer Defekte auf speziell präparierten Blechteilen in einer vorgegebenen Zeit erkennen und subjektiv nach ihrer Auffälligkeit bewerten sollten. Für die Schätzung der Detektionswahrscheinlichkeiten wurde das aus der probabilistischen Testtheorie bekannte Rasch-Modell verwendet. In der Studie wurden Bleche mit unterschiedlichen Glanzgraden untersucht und Defekte unterschiedlicher Größe und Ausprägung auf den Blechen zufällig verteilt. Zum Abgleich der Detektionsergebnisse der Testpersonen mit den deflektometrisch erfassten Eigenschaften der Defekte wird die psychometrische Funktion verwendet. Es wird gezeigt, dass sich die subjektiven Bewertungen der Studienteilnehmer mithilfe von Stevens Potenzgesetz ableiten lassen.

1 Einleitung

Im modernen Produktdesign haben spiegelnde Oberflächen eine wichtige Rolle. Neben ihrer Schutzfunktion ist die Beschaffenheit der Ober-

fläche maßgeblich für die Wahrnehmung eines Objekts verantwortlich. Beispiele hierfür sind lackierte Autokarosserien oder emaillierte Haushaltsgeräte. Während der Herstellung können Fehler auftreten, die erkannt und beurteilt werden müssen, um eine gleichbleibende Qualität zu gewährleisten. Studien zur Wahrnehmung von Defekten auf Oberflächen rücken daher immer mehr in den Fokus der klassischen Messtechnik. Da zunehmend automatische Inspektionssysteme die menschlichen Prüfer ablösen, ist es nötig, die bisher subjektiven Fehlermaße durch objektive Maße zu ersetzen. Diese sollten die Einschätzung eines menschlichen Prüfers bestmöglich nachbilden.

Insbesondere Fertigungstoleranzen, die eine reine Höhenabweichung zum Modell definieren, sind auf spiegelnden Oberflächen nicht geeignet. Im Gegensatz zu matten Oberflächen kann man in spiegelnden Oberflächen Teile der Umgebung wahrnehmen. Das virtuelle Bild der Umgebung, das auf der Oberfläche entsteht, hängt dabei von der Form und Reflektanz der Oberfläche ab. Da bereits kleine Änderungen in der Oberflächenkrümmung zu sichtbaren Verzerrungen des virtuellen Bildes der Umgebung führen, sind die Genauigkeitsanforderungen an solche Oberflächen höher als bei matten Oberflächen. Wie hoch diese Anforderungen für die Fehlermaße Tiefe, Ausdehnung und Krümmung eines Defekts sind und inwiefern sie von der Reflektanz der Oberfläche abhängen, wird in dieser Arbeit untersucht.

2 Deflektometrie

Bei der Inspektion spiegelnder Oberflächen treten im Vergleich zur Inspektion matter Oberflächen andere Problemstellungen auf. Zunächst lassen sich Messverfahren, die Muster auf die Oberfläche projizieren und die Oberfläche direkt beobachten, nur schwer anwenden. Hier finden deflektometrische Verfahren [1] ihre Anwendung. Zudem hat die Deflektometrie Vorteile, wenn es das Ziel ist, Fehler zu finden, die auch einem Menschen auffallen würden. Das Prinzip, mit dem die Oberfläche erfasst wird, ähnelt dem, wie auch der Mensch spiegelnde Oberflächen wahrnimmt. Statt die Oberfläche direkt zu beobachten, werden die Verzerrungen von Mustern, die sich in der Oberfläche spiegeln, gemessen. Die Verzerrungen sind durch die Form der Oberfläche vorgegeben und ändern sich dort am stärksten, wo auch die Krümmung der

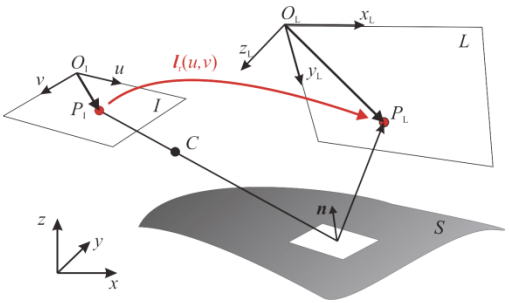


Abbildung 14.1: Prinzip der deflektometrischen Registrierung aus [1].

Oberfläche besonders groß ist. Diese Krümmungsinformationen sind im Normalenfeld der Oberfläche enthalten, welches von der Deflektometrie direkt erfasst wird. Dazu wird ein Messaufbau wie in Abbildung 14.1 dargestellt, bestehend aus einer Kamera mit der Bildebene I, der spiegelnden Oberfläche S als Prüfobjekt und dem Schirm L verwendet. Auf den Schirm werden sinusförmige Streifenmuster in horizontaler und vertikaler Richtung projiziert. Die Kamera ist so positioniert, dass die Muster auf dem Schirm über die spiegelnde Oberfläche beobachtet werden können. Durch die Abbildung über die Oberfläche werden die dargestellten Muster verzerrt. Anhand der Beobachtung einer ganzen Mustersequenz können die Schirmpunkte eindeutig den Kamerasichtstrahlen zugeordnet werden:

$$l : P_I \rightarrow P_L,\ l[u,v] = (x_L, y_L).$$

Diese Abbildung wird als deflektometrische Registrierung bezeichnet. Sie enthält bereits wesentliche Informationen über die Oberfläche, aus denen sich das Normalenfeld gewinnen lässt. Ohne Kenntnis der genauen Lage der Oberfläche lassen sich jedoch keine eindeutigen Oberflächennormalen angeben bzw. lässt sich keine eindeutige Rekonstruktion der Oberfläche erstellen [2].

3 Durchführung der Studie

Für die Studie wurden insgesamt 30 Mitarbeiter und Studenten am Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung

befragt. Wie im Überblick über das Teilnehmerfeld in Abbildung 14.2 zu sehen, ist die Testgruppe nicht repräsentativ für die Gesamtbevölkerung, gibt aber bereits Anhaltspunkte für eine Bewertung. Um Fremdeinflüsse weitestgehend auszuschließen, wurde die Studie in einem leeren Büroraum mit weiß gestrichenen Wänden, vollständig abgedunkeltem Fenster und geschlossener Tür durchgeführt. Die Studienteilnehmer saßen während ihrer Aufgabe an einem Tisch, auf dem ein 24"Monitor im Abstand von 20mm bis 110mm schräg im Winkel von 30° zur Tischplatte installiert wurde (siehe Abbildung 14.3). Auf dem Monitor wurde ein sinusförmiges Streifenmuster mit 50 Streifen angezeigt. Auf die Tischplatte unterhalb des Monitors wurde ein weißes Blatt mit aufgezeichnetem 10×10 - Raster gelegt, um es den Teilnehmern zu erleichtern, absolute Positionen auf dem Testblech bestimmen zu können. In die Mitte des Rasters wurden dann die 300mm $\times$ 300mm großen Testbleche gelegt. Die Testbleche wurden aus 1mm starkem Stahlblech gefertigt, schwarz lackiert und dann auf die Glanzgrade (nach DIN 67530) 60GE, 70GE, 80GE und 95GE poliert. Auf allen Testblechen wurden jeweils 8 Fehler (siehe Tabelle 14.1) an zufälligen Positionen (siehe Abbildung 14.4) aufgebracht. Zum Aufbringen der Fehler wurden die Testbleche in eine spezielle Halterung eingespannt und dann lokal mithilfe einer Gewindestange mit einem aufgesetzten Holzdübel verformt. Neben dem Monitor wurde der jeweils aktuelle Fragebogen abgelegt. Dazu wurde neben dem Monitor eine kleine Tischlampe mit 42W-Leuchtmittel angebracht, um in dem ansonsten dunklen Raum lesen zu können.

Defekt	1	2	3	4	5	7	8
Höhe	6mm	5mm	4mm	2mm	1mm	0,5mm	0,5mm

Tabelle 14.1: Höhe der auf den Testblechen aufgebrachten Defekte.

Der Testablauf war wie folgt: Zunächst wurden den Testpersonen allgemeine Testanweisungen und Fragebogen mit allgemeinen Fragen zur Person vorgelegt. Nachdem die allgemeinen Fragen beantwortet waren, bekamen die Testpersonen nacheinander vier Testbleche mit den Glanzgraden 60GE, 70GE, 80GE und 95GE vorgelegt und jeweils einen dazugehörigen Fragebogen mit zwei Aufgaben. Die erste Aufgabe bestand darin, alle Defekte, die auf dem Blech erkannt wordene waren, fortlaufend nummeriert in ein 10×10 - Raster einzutragen. Zum Be-

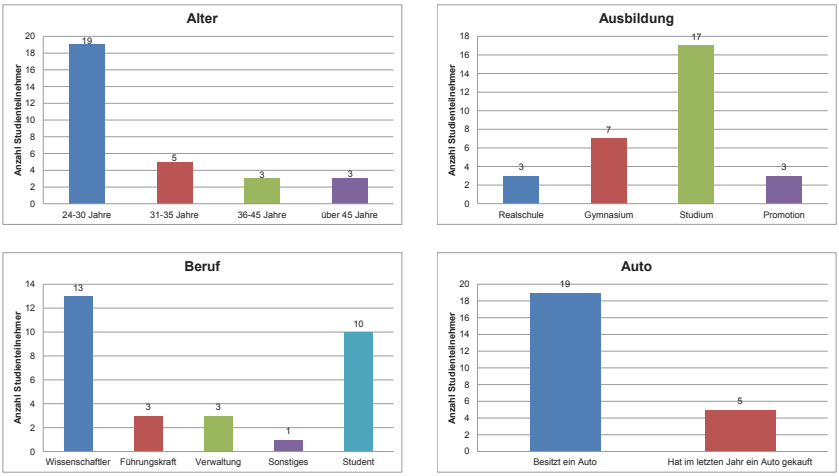


Abbildung 14.2: Überblick über die 30 Studienteilnehmer.

arbeiten dieser Aufgabe hatten die Testpersonen jeweils maximal eine Minute Zeit. Bei der zweiten Aufgabe sollte nun jedem erkannten Defekt eine Bewertung auf einer Skala (1–6, einfach zu finden – schwer zu finden) zugeordnet werden. Für die zweite Aufgabe gab es keine Zeitbeschränkung. Der gesamte Test dauerte etwa 20 Minuten pro Person.

4 Methoden

Um die Detektionswahrscheinlichkeiten der einzelnen Defekte zu bestimmen, wurde das Rasch-Modell [3] verwendet. Das Rasch-Modell ist ein Modell der probabilistischen Testtheorie und unterscheidet zwischen der (im Allgemeinen nicht direkt beobachtbaren) Fähigkeit einer Person, bestimmte Aufgaben zu lösen, und der Schwierigkeit der jeweiligen Aufgaben. Damit ist es möglich, die Lösungswahrscheinlichkeit der Aufgabe für eine durchschnittliche Person zu schätzen, unabhängig davon, wie genau das Teilnehmerfeld zusammengesetzt ist. Zur Schätzung der Rasch-Parameter wurde die bedingte Maximum-Likelihood-Methode verwendet (siehe [4], Kapitel 10 und [5]). Da für

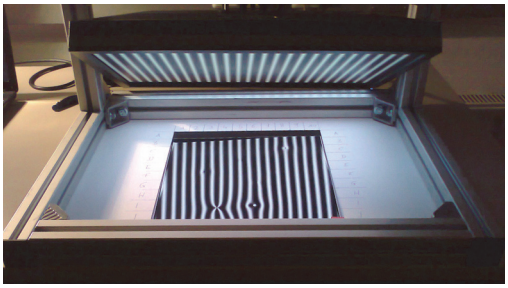
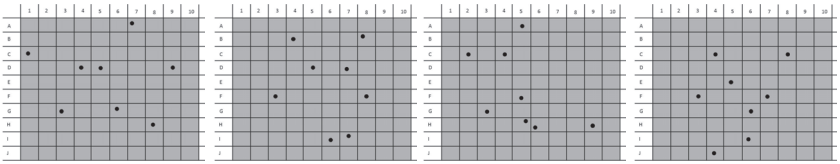


Abbildung 14.3: Testaufbau zur Wahrnehmungsstudie.



(a) 60GE (b) 70GE (c) 80GE (d) 95GE

Abbildung 14.4: Verteilung der Fehler auf den Testblechen.

Aufgaben, die von keinen bzw. allen Testpersonen gelöst werden, die Aufgabenschwierigkeit nur in eine Richtung abgeschätzt werden kann, wurde die Wahrscheinlichkeit für die Lösung einer solchen Aufgabe 0 bzw. 1 gesetzt.

Für die Beschreibung des Zusammenhangs der gemessenen Höhe, Fläche bzw. Krümmung eines Defekts und der Wahrscheinlichkeit, dass ein Studienteilnehmer diesen entdeckt, wurde die psychometrische Funktion [6] verwendet:

$$F(x) = \gamma + (1 - \lambda - \gamma) f(x) \text{ mit } f(x) = 1 - e^{-(\beta x)^{\alpha}}.$$

Dabei ist γ die Wahrscheinlichkeit dafür, eine richtige Lösung zu raten, und λ die Wahrscheinlichkeit dafür, auch bei einer sicheren Lösung der Aufgabe einen Fehler zu machen. Da die besonders einfachen Fehler von allen Testpersonen gefunden wurden, wurde hier $\lambda = 0$ angenommen. Die Wahrscheinlichkeit, einen der 8 Fehler auf dem 10×10 - Raster zu erraten, wird im Mittel mit $4/100$ abgeschätzt, daher wurde γ entsprechend gewählt. Als Wahrscheinlichkeitsverteilung $f(x)$ wurde hier

die in der Literatur häufig verwendete Weibull-Verteilung [6] mit $\alpha = 2$ angenommen, da eine gleichzeitige zuverlässige Schätzung von α und β aufgrund der wenigen Datenpunkte nicht möglich ist.

Um aus den gemessenen Größen Höhe h, Fläche A und Krümmung κ eines Defekts die subjektive Bewertung der Studienteilnehmer abzuleiten, wurde die Stevens'sche Potenzfunktion [7] für Reiz R und Empfindung E verwendet:

$$E = k(R - R_0)^n.$$

Sie verallgemeinert das Weber-Fechner'sche Gesetz, wonach der empfundene Sinneseindruck proportional zum Logarithmus der Stärke des physikalischen Reizes ist, was oftmals nur für kleine Intensitätsbereiche gültig ist. Da die Studienteilnehmer die Bewertungsskala 1–6 unterschiedlich nutzten, mussten die Bewertungen der Defekte zunächst normalisiert werden, um diese zwischen den Studienteilnehmern vergleichbar zu machen. Dafür wurden die Bewertungen auf eine Skala von $0 =$ „sehr schwer zu sehen“ und $1 =$ „sehr einfach zu sehen“ normalisiert. Zudem konnte in den Daten beobachtet werden, dass die Teilnehmer die Skala nicht gleichbleibend über die verschiedenen Bleche hinweg verwendet haben. So haben sie für Fehler gleicher Größe auf den unterschiedlichen Glanzgraden die gleichen Bewertungen vergeben, obwohl die Detektionswahrscheinlichkeiten zeigen, dass die Fehler auf den stärker spiegelnden Oberflächen einfacher zu sehen sind. Ein Vergleich der Glanzgrade aufgrund der Bewertungen ist somit nicht mehr möglich. Zudem wurde angenommen, dass ein Fehler, der von einem Teilnehmer nicht gefunden wurde und in Folge dessen auch nicht bewertet wurde, von dieser Person als maximal schwer bewertet worden wäre. Die Bewertungen von 2 Teilnehmern waren entgegengesetzt zu denen aller anderen Teilnehmer, so dass hier von einem Fehler bei der Beantwortung ausgegangen werden kann. Die entsprechenden Bewertungen wurden aus der Auswertung ausgeschlossen.

Die physikalischen Größen wurden deflektometrisch bestimmt, wobei hier nicht kalibrierte Messungen verwendet wurden. Somit sind die Werte für Fläche und Krümmung lediglich proportional zu den wahren Werten, aber für zwischen den verschiedenen Blechen vergleichbar. Die Fläche der Defekte ist außerdem von dem Schwellwert abhängig, ab dem eine Abweichung von der Ebene als Defekt angesehen wird.

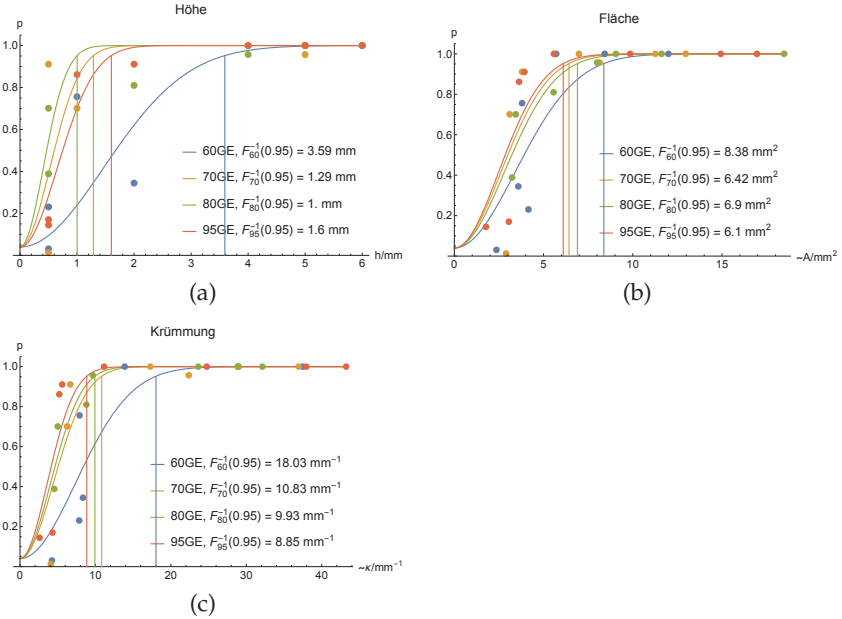


Abbildung 14.5: Detektionswahrscheinlichkeiten in Abhängigkeit von der Höhe, der Fläche und der Krümmung (jeweils unkalibriert).

Die Höhe der Defekte wurde über die Einschraubtiefe beim Auftragen der Defekte bestimmt.

5 Ergebnisse

In der Studie zeigten sich zunächst verschiedene Probleme in den Daten, die die Ergebnisse beeinflussten. Sobald zwei Defekte auf der Oberfläche sehr nah beeinander lagen, fiel es den Testpersonen deutlich schwerer, beide Defekte getrennt voneinander zu erkennen (z.B. Defekte auf H5/H6 in Abbildung 14.4c). Die entsprechenden Ergebnisse mussten aus der Auswertung entfernt werden, um das Ergebnis nicht zu verfälschen. Außerdem setzte eine Art Lernprozess bei der Bearbeitung der Aufgaben ein. Mit jedem Blech wurde die Zeit kürzer, die die

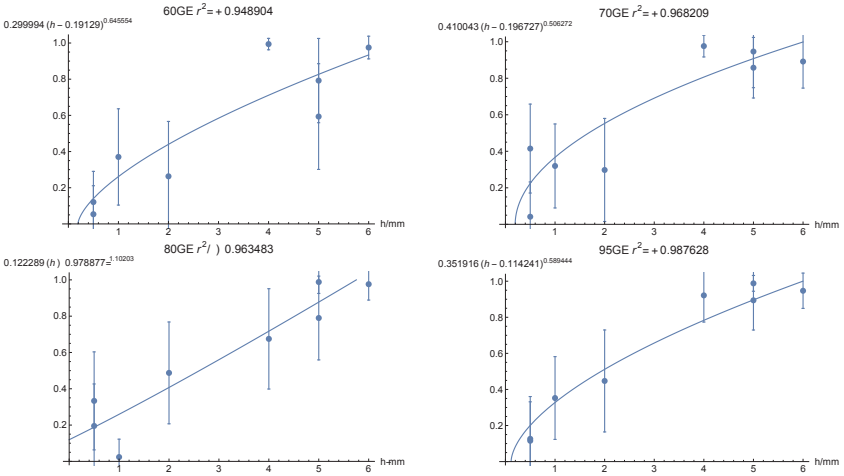


Abbildung 14.6: Normalisierte Bewertung und Standardabweichung in Abhängigkeit von der Höhe (unkalibriert) für die 4 Testbleche.

Personen für das Suchen und Dokumentieren der Fehler benötigten. Da jedes Mal mit dem Blech 60GE angefangen wurde, kann davon ausgegangen werden, dass die Detektionsleistungen hier etwas zu niedrig eingeschätzt werden.

In den Grafiken in Abbildung 14.5 sind die Detektionswahrscheinlichkeiten der Defekte in Abhängigkeit von Höhe, Fläche und Krümmung der Defekte dargestellt. Zudem wurde die psychometrische Funktion an die Ergebnisse angepasst und der Wert der jeweiligen Größe bestimmt, ab dem die Erkennungswahrscheinlichkeit $95\,\%$ beträgt. Es ist zu beobachten, dass die Erkennungswahrscheinlichkeit für das schwach spiegelnde Blech (60GE) in allen drei Fällen stark abnimmt. Für die stärker spiegelnden Bleche zeigt sich kein eindeutiges Bild, da die Unterschiede recht gering sind und sich die Reihenfolge in den Abbildungen a, b und c ändert. Das erwartete Ergebnis, dass Defekte auf stärker spiegelnden Blechen eher gefunden werden, zeigt sich einzig in Abhängigkeit von der Krümmung (Abbildung 14.5c). Die exakten Werte für die Sichtbarkeit sollten mit Vorsicht behandelt werden, da die Modellparameter aufgrund der wenigen Abtastpunkte (8 Fehler pro Blech)

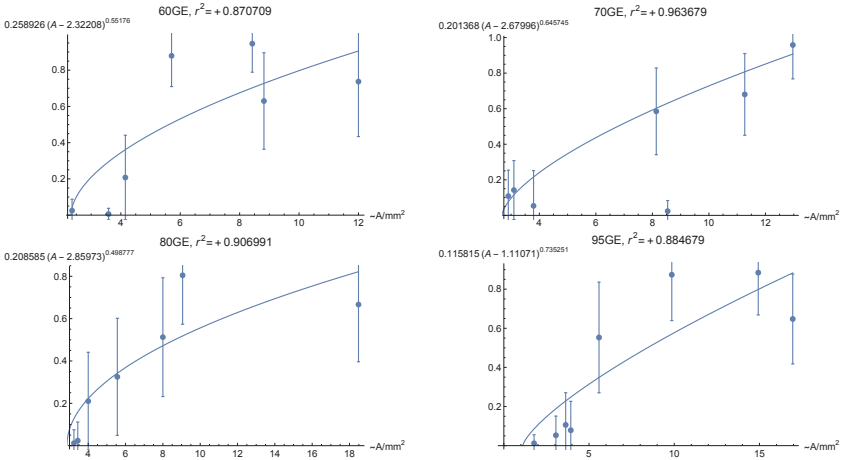


Abbildung 14.7: Normalisierte Bewertung und Standardabweichung in Abhängigkeit von der Fläche (unkalibriert) für die 4 Testbleche.

statistisch nicht signifikant sind.

Die Abbildungen 14.6, 14.7 und 14.8 zeigen die normalisierten Bewertungen der Defekte durch die Testteilnehmer in Abhängigkeit von den physikalischen Messgrößen und dem Glanzgrad des Blechs. Zudem ist die Standardabweichung für die Bewertung jedes Defekts gegeben. Darüber hinaus ist das beste angepasste Modell (Stevens'sche Potenzfunktion) dargestellt und dessen Bestimmtheitsmaß[4] r^2 gegeben. Die geschätzten Exponenten n des Modells nehmen Werte kleiner als 1 an. Das bedeutet, dass Änderungen in der Defektausprägung gerade im Grenzbereich der Wahrnehmbarkeit der Defekte stärker wahrgenommen werden. Aufgrund der wenigen Datenpunkte und der hohen Varianz in den Bewertungen ist die statistische Signifikanz der Modellparameter allerdings zu gering, um damit verlässliche Vorhersagen zu treffen. Es zeigt sich jedoch, dass sich die Bewertungen mithilfe von Stevens Potenzfunktion aus den messbaren Größen Höhe, Ausdehnung und Krümmung ableiten lassen. Die Bestimmtheitsmaße für die Modelle, die den Zusammenhang von Höhe bzw. Krümmung eines Fehlers

[4] Kein Zusammenhang von Modell und Daten besteht für $r^2 = 0$. Ein perfekter Zusammenhang von Modell und Daten liegt für $r^2 = 1$ vor.

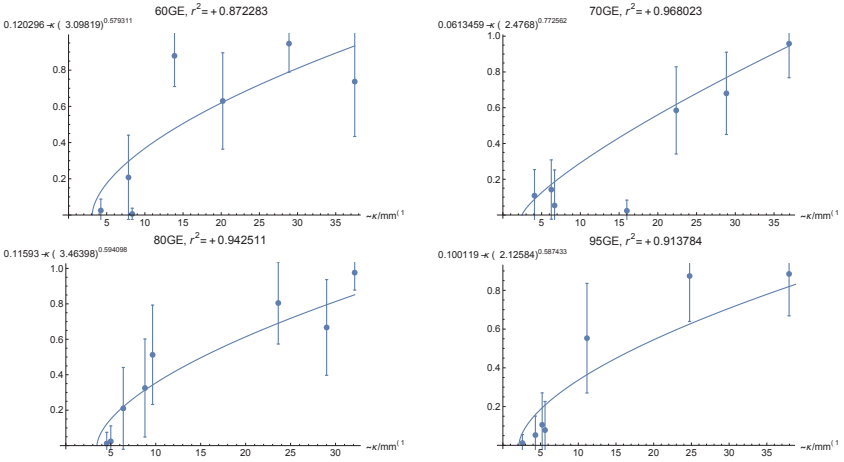


Abbildung 14.8: Normalisierte Bewertung und Standardabweichung in Abhängigkeit von der Krümmung (unkalibriert) für die 4 Testbleche.

und dessen Bewertung darstellen, sind besser als die Bestimmheitsmaße für die Modelle, die den Zusammenhang von Fläche und Bewertung darstellen.

6 Zusammenfassung

Unter Verwendung mehrerer Standardverfahren aus der Psychometrie und den empirischen Sozialstudien wurde der Zusammenhang deflektometrisch messbarer Größen spiegelnder Oberflächen und der Wahrnehmung des Menschen hergestellt. Dabei stellten sich die Höhe und die Krümmung als geeignete Maße für die Detektionswahrscheinlichkeit und die subjektive Bewertung von Fehlern auf spiegelnden Oberflächen heraus. Die Vermutung, dass die Krümmung besonders gut zur Vorhersage der Sichtbarkeit geeignet ist, konnte nicht eindeutig bestätigt werden. Für die Vorhersage von Detektionswahrscheinlichkeiten und Bewertungen allein aus Messgrößen der Deflektometrie und dem Glanzgrad der Oberfläche müssen jedoch deutlich mehr Defekte untersucht werden.

In der Studie hat sich eine Reihe von Verbesserungsmöglichkeiten gezeigt: Wird die Studie wie in diesem Artikel beschrieben durchgeführt, sollte die Reihenfolge der Bleche für jeden Studienteilnehmer anders sein, um den systematischen Einfluss von Lerneffekten auszuschließen. Außerdem sollte jeder Teilnehmer zunächst die Möglichkeit haben, auf einem Trainingsblech die Defektbewertung zu üben, bevor das erste Blech bewertet wird. Um auszuschließen, dass sich die Fehler gegenseitig beeinflussen, muss ein Mindestabstand zwischen den einzelnen Fehlern sichergestellt sein. Für die Auswertung der subjektiven Bewertungen wäre eine gleichbleibend, objektive Referenzskala hilfreich, z.B. in Form eines Referenzbleches. Außerdem kann man die genauen Wahrnehmungsschwellen schneller bestimmen, indem man einen adaptiven Test, d.h. einen Test mit Defekten, die nah an der Wahrnehmungsschwelle der jeweiligen Person liegen, durchführt.

7 Danksagung

Wir danken Benjamin Krüger, der diese Studie im Rahmen seiner Bachelorarbeit durchgeführt hat. Diese Veröffentlichung ist im Rahmen des Projekts MID-Wave entstanden. Dieses Projekt wird finanziert durch die Baden-Württemberg Stiftung gGmbH.

Literatur

1. S. Werling, „Deflektometrie zur automatischen Sichtprüfung und Rekonstruktion spiegelnder Oberflächen“, Dissertation, Karlsruher Institut für Technologie, 2011.
2. J. Balzer, „Regularisierung des Deflektometrieproblems – Grundlagen und Anwendung“, Dissertation, Universität Karlsruhe (TH), Universitätsverlag Karlsruhe, 2008.
3. G. Rasch, „On general laws and the meaning of measurement in psychology“, in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, Vol. 4, 1961, S. 321–333.
4. H. Moosbrugger und A. Kelava, *Testtheorie und Fragebogenkonstruktion*. Springer Berlin Heidelberg, 2012.
5. P. Mair, R. Hatzinger und M. Maier, „Extended rasch modeling: The r package erm“, Wirtschaftsuniversität Wien, Tech. Rep., 2009.

6. F. A. Wichmann und N. J. Hill, „The psychometric function: I. fitting, sampling, and goodness of fit“, *Perception & psychophysics*, Vol. 63, S. 1293–1313, 2001.
7. S. S. Stevens, „On the psychophysical law.“ *Psychological review*, Vol. 64, S. 153–181, 1957.

Bildverarbeitungsbasierte Quantifizierung der Konfluenz von Stammzellkolonien zur Prozesssteuerung in einer Bioproduktionsanlage

Friedrich Schenk[1], Christian Kowalski[1] und Robert Schmitt[1,2]

[1] Fraunhofer-Institut für Produktionstechnologie IPT,
Steinbachstraße 17, 52074 Aachen
[2] RWTH Aachen, Werkzeugmaschinenlaboratorium (WZL),
Lehrstuhl für Fertigungsmesstechnik und Qualitätsmanagement,
Steinbachstraße 19, 52074 Aachen

Zusammenfassung Zur Prozesssteuerung in einer Bioproduktionsanlage wurde ein Bildverarbeitungsalgorithmus zur Quantifizierung der Konfluenz (Besiedlungsdichte) von Stammzellkolonien entwickelt. Anhand der ermittelten Konfluenzwerte wird der richtige Zeitpunkt zur Passagierung der Zellkolonien abgeleitet. Dies ist notwendig, um ideale Wachstumsbedingungen der Zellkultur zu gewährleisten. Der Algorithmus sorgt durch eine Kombination von schwellwertbasierten Verfahren mit morphologischen Operationen für eine robuste Segmentierung auch bei Beleuchtungsinhomogenitäten. Die Algorithmen wurden speziell für die Prozessierung großer Bilddaten ausgelegt.

1 Einleitung

In den letzten Jahren hat die Automatisierungstechnik auch im Bereich der Zellkultur immer stärker Einzug gehalten. Gerade wenn Zellen im großen Maßstab kultiviert werden sollen, lohnt es sich, manuelle Tätigkeiten automatisiert durchzuführen. Dabei muss die menschliche Expertise durch eine intelligente Mess- und Steuerungstechnik ersetzt werden. Die Automatisierung erlaubt nicht nur einen höheren Durchsatz, sondern sorgt aufgrund standardisiert ablaufender Prozesse für eine gleichmäßigere und bessere Qualität der produzierten Zellen. Im

Rahmen des Forschungsprojektes *StemCellFactory* wurde eine Anlage zur vollautomatisierten Produktion induziert pluripotenter Stammzellen (iPS Zellen) entwickelt [1].

1.1 iPS-Zellproduktion

Während pluripotente embryonale Stammzellen aus frühen menschlichen Embryonen gewonnen werden und die Gewinnung somit erstens beschränkt und zweitens ethisch und moralisch bedenklich ist, können iPS Zellen ethisch unbedenklich erzeugt werden. Das Verfahren wurde im Jahr 2006 durch eine Gruppe um Yamanaka entwickelt und basiert auf der Einschleusung von vier Genen (c-myc, Klf4, Sox2, Oct4) in menschliche Zellen, was zu einer Reprogrammierung führt [2]. Somit ist es möglich z. B. Hautzellen in eine Art embryonalen Zustand zurückzuversetzen. Aus solchen patientenspezifischen Stammzellen können verschiedenste Zellprodukte abgeleitet werden. Von besonderer Bedeutung ist hierbei die Möglichkeit, diese Zellen zur patientenspezifischen Wirkstoffentwicklung zu nutzen. Voraussetzung für eine Wirkstofftestung im großen Maßstab ist die Generierung einer Vielzahl an Zellen in gleichbleibender Qualität. Die Herstellung und Kultivierung von iPS Stammzellen ist allerdings ein zeit- und personalaufwändiger Vorgang. Zudem ist eine schwankende Zellqualität bei manueller Herstellung nicht zu vermeiden, da insbesondere „Einflüsse wie Erfahrung, Geschicklichkeit als auch Verfassung des Personals [...] direkte Auswirkungen auf die Qualität der erzeugten iPS-Zellen“ [3] haben. Daher ist die Automatisierung des kompletten Herstellungsprozesses von iPS Stammzellen sehr attraktiv, um die Quantität und Qualität der produzierten Zellen zu erhöhen.

1.2 Aspekte der Zellkultur

Da biologische Prozesse sehr variabel sind und zum Teil unvorhersehbar ablaufen, ist eine präzise Kontrolle und Überwachung der Kulturen notwendig. Dies geschieht oftmals bildgestützt unter Einsatz von Bildverarbeitungsalgorithmen. Eine flexible Prozesssteuerung sorgt auf Basis der Messwerte für eine optimierte Zellkultivierung.

Als Zellkultur wird das Kultivieren von tierischen oder pflanzlichen Zellen *in vitro*, also außerhalb eines Lebewesens, in einem Nährmedium

verstanden [4]. Es existieren unterschiedliche Arten von Zellkulturen. Unter einer Primärkultur versteht man solch eine Zellkultur, deren Lebensdauer *in vitro* begrenzt ist und Gewebe oder Organe als Ausgangsmaterial nutzt. Durch die Subkultivierung dieser Primärkultur entsteht eine Sekundärkultur. Wird diese Sekundärkultur erneut passagiert, spricht man von einer Zelllinie [5]. Weiterhin wird bei der Primärkultur zwischen adhärenter Zellkultur und Suspensionskultur unterschieden. Während adhärente Zellen auf einem Substrat haften und in einer einlagig zusammenhängenden Schicht (Monolayer) wachsen, reifen Suspensionszellen als Einzelzellen oder Zellklümpchen in Suspension heran [4]. Es existieren spezielle Zellkulturgefäße, beispielsweise Mikrotiterplatten (MTPs), in denen adhärente Zellen in einzelnen Bereichen (Wells) auf dem Plattenboden haften. Mit den Nährstoffen des umgebenden Zellkulturmediums versorgt, wachsen die Zellen in Kolonieverbünden immer großflächiger. Während dieses Wachstumsprozesses nimmt die Konzentration an Nährstoffen im Medium stetig ab, während Stoffwechselendprodukte zunehmen. Daher ist ein regelmäßiger Wechsel des Zellkulturmediums essentiell für das ideale Wachstum und zur Vitalitätserhaltung der Zellkultur [5].

Darüber hinaus ist es sowohl bei adhärenten als auch bei Suspensionszellen notwendig, eine Subkultivierung durchzuführen. Das geschieht durch den Vorgang des Passagierens, der Vereinzelung und Umsetzung der Zellen und damit Bereitstellung zusätzlichen Raums für weiteres Wachstum. Dabei ist es entscheidend, die Zellen rechtzeitig zu passagieren. Dies ist notwendig, da mit steigender Zelldichte das Nährstoffangebot sowie der pH-Wert des Mediums sinken und sich die Wachstumsbedingungen verschlechtern. Zudem entsteht ein Selektionsdruck, der solche Zellen begünstigt, die selbst mit ungünstigen Bedingungen noch zurechtkommen, wodurch sich die Zellpopulation auf die Dauer verändert. Darüber hinaus wird das Wachstum vieler Zellen unterbrochen, sobald diese infolge fehlender Fläche zwangsweise in Kontakt treten (Kontaktinhibition). Die Konfluenz (Besiedlungsdichte) des Zellrasens muss daher kontinuierlich überwacht werden, um ein zu dichtes Wachstum des Zell-Monolayers zu verhindern.

In der manuellen Zellkultur werden für die Bestimmung des Passagierzeitpunktes entweder Zellzählungen durchgeführt oder es wird nach Erfahrungswert entschieden, ob eine Zellkultur eine ausreichende Konfluenz aufweist, um passagiert zu werden. In der au-

tomatisierten Zellkultur sollte diese Kontrolle bildverarbeitungsbasiert anhand von Mikroskopaufnahmen durchgeführt werden. In dem Maßstab der *StemCellFactory* betrifft dies bei voller Auslastung täglich bis zu zweihundert Mikrotiterplatten, die mikroskopiert werden müssen. Um diesen Durchsatz bewältigen zu können, wurde eine spezielle Hochdurchsatz-Mikroskopielösung entwickelt [6]. An diese schließt sich die bildverarbeitungsgestützte Konfluenzquantifizierung der Stammzellkolonien an. Dabei dienen große Übersichtsaufnahmen, die durch Stitching zahlreicher mikroskopischer Einzelaufnahmen entstanden sind, als Ausgangsmaterial für den Bildverarbeitungsalgorithmus.

1.3 Bilddaten

Bei dem Mikroskop in der *StemCellFactory* handelt es sich um ein inverses Forschungsmikroskop vom Typ Nikon Ti-E, das im Durchlicht Hellfeld- und Phasenkontrastaufnahmen im hohen Durchsatz ermöglicht (siehe Abb. 15.1). Als Kamera dient eine 4/3″ sCMOS Kamera mit 2560 x 2160 Pixeln vom Typ pco.edge. Durch die Aufnahme

Abbildung 15.1: Hochdurchsatzfähiges Mikroskop in der *StemCellFactory*.

der Bilder im Phasenkontrast sind die fast durchsichtigen Stammzellen gut sichtbar und heben sich vom Hintergrund merklich ab, was eine Segmentierung vereinfacht. Bei der Phasenkontrastbildgebung von

Mikrotiterplatten tritt jedoch ein Artefakt, der sogenannte Randeffekt auf, der durch die Oberflächenspannung des Zellkulturmediums und die Adhäsionskräfte zum Innenrand des Wells verursacht wird. Die konkav gekrümmte Flüssigkeitsoberfläche sorgt für eine Lichtbrechung und eine Verschiebung des Ringbilds der Ringblende. Je stärker die Krümmung, desto weniger liegen das Bild der Ringblende und der Phasenring des Objektivs aufeinander. Damit geht der Phasenkontrast im Randbereich verloren und es kommt zu einer Überblendung. Die im Hellfeld aufgenommenen Bilder besitzen keinen sichtbaren Randeffekt, allerdings verfügen diese Bilder auch über einen weitaus geringeren Kontrast, sodass sich die Stammzellkolonien nur schwach vom Hintergrund abheben.

2 Quantifizierung der Konfluenz

µZellzählungen zur Bestimmung der Konfluenz sind sehr zeitintensiv. Andererseits ist die Passage zu festgelegten Zeitpunkten bei der hohen Variablität des Zellwachstums ungünstig. Folglich spart die Quantifizierung der Konfluenz durch bildverarbeitungsbasierte Algorithmen nicht nur Zeit und Personal, sie liefert bei robustem Algorithmus außerdem ein besseres Ergebnis.

Für die Quantifizierung der Konfluenz müssen zunächst Bilddaten in ausreichender Qualität vorliegen. Dies betrifft sowohl die Schärfe als auch die homogene Ausleuchtung der Aufnahmen. Je hochwertiger das Bildmaterial, desto robuster kann anschließend der Bildverarbeitungsalgorithmus funktionieren. In der *StemCellFactory* reifen die zu untersuchenden Stammzellkolonien in einem automatisierten Inkubator (LiCONiC Instruments) heran. Sämtliche Prozessschritte werden außerhalb des Inkubators in der eigentlichen Anlage vorgenommen. Dazu entnimmt ein Reinraumroboter (Kuka) die MTPs und transportiert sie zur entsprechenden Station. Für die Konfluenzbestimmung wird die Mikrotiterplatte mit den zu untersuchenden Zellen zum High-Speed Mikroskop (modifiziertes Nikon Ti-E) transportiert und innerhalb kürzester Zeit mit der gewünschten Vergrößerung im Phasenkontrast digitalisiert. Die Einzelbilder werden zu einem Gesamtbild durch einen Stitching-Algorithmus zusammengefügt und nehmen damit ein großes Datenvolumen ein (mehrere Gigabytes pro Platte). Nachdem

die Bilddaten vorliegen, wird die Quantifizierung der Konfluenz durch den Bildverarbeitungsalgorithmus berechnet. Dabei werden die Bilddaten vor der Auswertung durch den Algorithmus herunterskaliert. Die reine Konfluenzbestimmung benötigt nicht so hochaufgelöste Aufnahmen und läuft mit geringeren Datenmengen effizienter und schneller. Abhängig vom Ergebnis der Konfluenzbestimmung wird die Mikrotiterplatte mit den iPS-Stammzellkolonien passagiert oder bei einer zu geringen Besiedlungsdichte zur weiteren Reifung zurück in den Inkubator transportiert.

2.1 Algorithmus

Der Bildverarbeitungsalgorithmus verfügt über folgende Module:

- Metadaten auslesen und Einstellungen laden
- Bilddaten komprimieren
- Region of Interest (ROI) setzen
- Segmentierung durchführen
- Partikelanalyse und Konfluenzbestimmung

Alle Aufnahmeeinstellungen, insbesondere die Vergrößerung und der Mikroskopiermodus, werden aus Metadatensätzen ausgelesen und beeinflussen die nachfolgenden Verarbeitungsschritte. Zur Reduktion der Bildgröße wird eine Skalierung vorgenommen. Der Skalierungsfaktor ist dabei von der Aufnahmevergrößerung abhängig und ist bei einem 4x-Objektiv zu 0,3 gesetzt. Damit wird ein zusammen gestitchtes Einzelbild einer 6-Well Mikrotiterplatte von 25600 x 25920 Pixeln (10 x 12 Bilder) auf 7680 x 7776 Pixel herunterskaliert. Dies hat sich als hinreichend hohe Auflösung heraus gestellt.

Da ausschließlich Mikrotiterplatten mit runden Wells benutzt werden, die Bilddaten aufgrund des rechteckigen Bildsensors aber rechteckig vorliegen und mehr als das eigentliche Well beinhalten, muss eine kreisförmige Region of Interest (ROI) gesetzt werden, die den Grenzen der Wells entspricht.

Der Segmentierungsalgorithmus beginnt mit einer Kantendetektion mittels des Prewitt-Operators. Dieser Gradientenfilter nutzt folgende

3 x 3 Filterkerne P_x und P_y [7]:

$$P_x = \begin{pmatrix} 0 & 0 & 0 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \end{pmatrix} \qquad P_y = \begin{pmatrix} 0 & -1 & 1 \\ 0 & -1 & 1 \\ 0 & -1 & 1 \end{pmatrix} \tag{15.1}$$

Im Anschluss an die Kantendetektion wird der Bildkontrast durch eine Histogrammbegrenzung und anschließende Histogrammspreizung erhöht. Alternativ können auch die Grauwerte des Bildes mit einem Faktor multipliziert, damit sie sich stärker voneinander abheben wie Abbildung 15.3 zeigt. Diese Operation hat auf höhere Grauwerte einen stärkeren Einfluss als auf niedrigere Grauwerte. Anschließend findet über ein Schwellenwertverfahren mit definierter Schwelle eine Binarisierung des Bildes statt. Der Schwellenwert wird dabei entweder nach Erfahrungswerten manuell gesetzt oder automatisiert anhand der Auswertung des durchschnittlichen Grauwertes eines Linienprofils innerhalb der ROI. Da die Zellkolonien im Phasenkontrast stets heller erscheinen als der Hintergrund (siehe Abb. 15.2), hat ein Linienprofil, das sich über Kolonien und Hintergrundbereiche erstreckt, einen durchschnittlichen Grauwert, der sich gut als automatisch bestimmte Segmentierungsschwelle eignet.

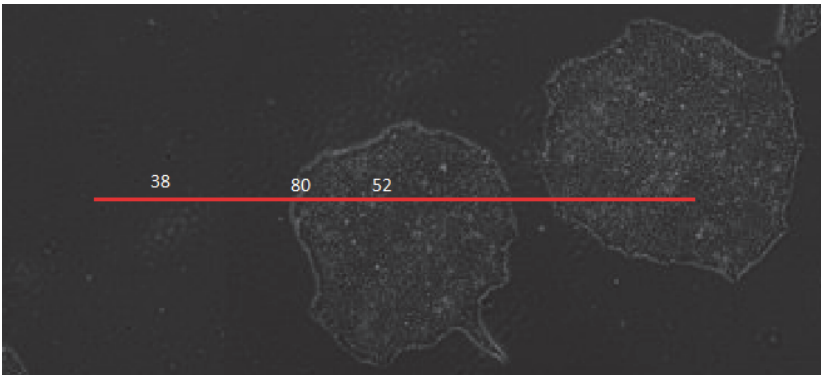


Abbildung 15.2: Automatisierte Bestimmung des Schwellwertes durch Auswertung des durchschnittlichen Grauwertes eines Linienprofils.

Nachdem das Binärbild vorliegt, wird das Segmentierungsergebnis durch morphologische Filter nachbearbeitet. Durch eine zweimalige

Erosion mit einem 3 x 3 Strukturelement werden kleinste Partikel entfernt, die nicht zu den Stammzellkolonien gehören, sondern Segmentierungsartefakte darstellen. Im letzten Schritt werden kleine Löcher im Inneren der segmentierten Bereiche durch die Closing-Operation geschlossen, ohne das Objekt selbst signifikant zu modifizieren. Beim Closing handelt es sich um die Abfolge einer Dilatation mit anschließender Erosion mit identischem Strukturelement, in unserem Fall mit der Dimension 7 x 7. Die Abfolge der genannten Operationen bildet den Segmentierungsalgorithmus.

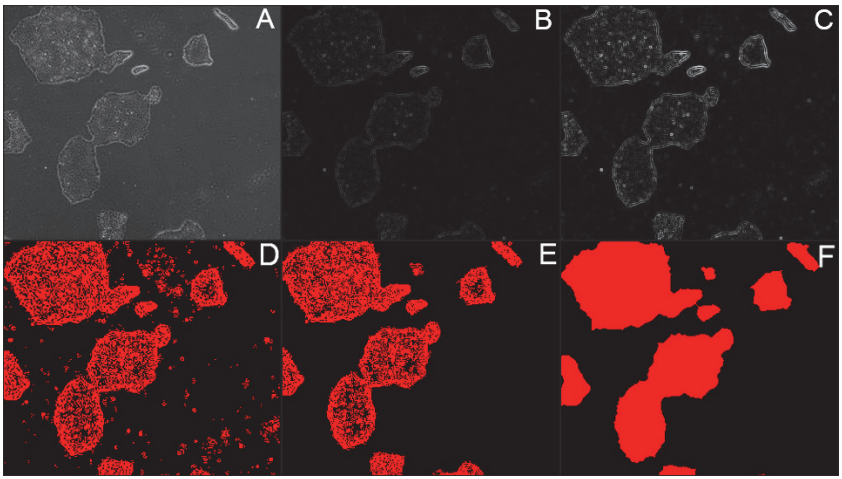


Abbildung 15.3: Ablauf der Segmentierung: (A) Originalbild; (B) Anwendung des Prewitt-Operators; (C) Kontrasterhöhung durch Multiplikation mit Konstante; (D) Schwellenwertverfahren; (E) Entfernen kleiner Objekte durch zweifache Erosion; (F) Schließen von Löchern durch Closing-Operation.

Das Verhältnis des segmentierten Bereichs zu der unsegmentierten Hintergrundfläche ergibt die Konfluenz.

$$\frac{\text{Summe der segmentierten Bereiche}}{\text{Pixelanzahl der ROI}} \cdot 100 = \text{Konfluenz [\%]} \quad (15.2)$$

Die segmentierten Bereiche werden anschließend einer weitergehenden Partikelanalyse unterzogen, in der beispielsweise kleinste und

größte Kolonien sowie durchschnittliche Koloniegrößen bestimmt werden. Weitere Statistiken wie die Anzahl an Kolonien einer gewissen Mindestgröße sind problemlos realisierbar.

2.2 Ergebnisse

Da die Segmentierung auf einer Kantendetektion beruht, sind scharfe, fokussierte Aufnahmen eine Grundvoraussetzung für eine exakte Konfluenzbestimmung. Wie Abb. 15.4 zeigt, führt nur das richtig fokussierte Bild C zu einer korrekten Segmentierung. Bei unscharfen Kanten fällt der als Zellkolonie detektierte Bereich (F,G, I, J) zu groß aus.

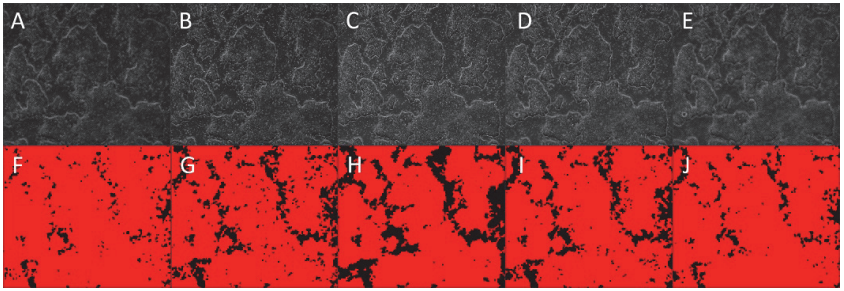


Abbildung 15.4: Segmentierungsergebnis bei unterschiedlichen Fokuslagen Defokussierung A-E: -100 µm, -50 µm, 0 µm, +50 µm, +100 µm

Bei entsprechendem Bildmaterial liefert der Algorithmus zur Konfluenzdetektion sehr robuste Segmentierungsergebnisse. Selbst in den Randbereichen der Wells, in denen die Zellen infolge des Randeffekts überbelichtet und kontrastarm erscheinen, können die Zellen zuverlässig segmentiert werden. Auch gegenüber anderen Artefakten wie beispielsweise Beschlag am Deckel, der zu Abschattungen führt (vgl. Abb. 15.5 B), ist der Algorithmus relativ unempfindlich.

Abbildung 15.6 zeigt drei verschiedene Wells einer 24-Well Mikrotiterplatte zu zwei unterschiedlichen Zeitpunkten. Bei Teil A-C der Abbildung handelt es sich um iPS-Zellen am vierten Tag nach der Passage, Teil D-F zeigt die gleichen Wells nach drei weiteren Tagen. Es ist zu erkennen, dass alle drei Wells an Tag 7 einen bis zu zehnmal höheren Konfluenzwert aufweisen. Das Segmentierungsergebnis ist jeweils als

Overlay in den Wells abgebildet. Er wird deutlich, dass die Zellkolonien innerhalb der Region of Interest zum größten Teil korrekt detektiert und quantifiziert werden konnten.

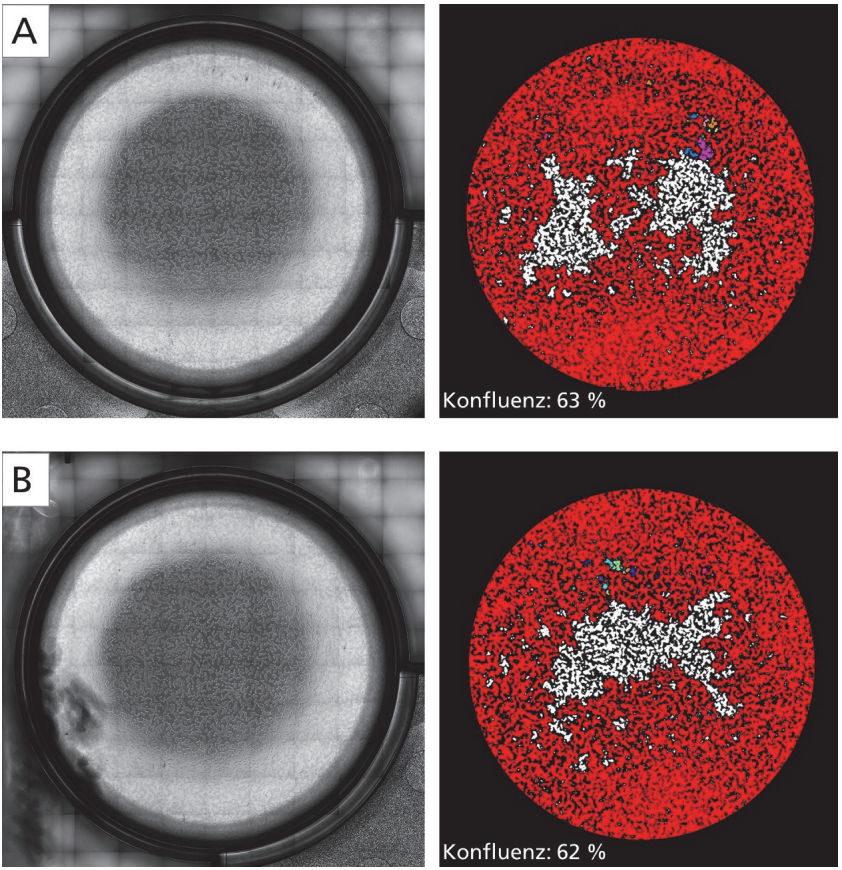


Abbildung 15.5: Bestimmung der Konfluenz einer 6-Well MTP (Phasenkontrast): (A) Artefaktfreie Aufnahme; (B) Artefakt durch Beschlag an Deckel der Mikrotiterplatte.

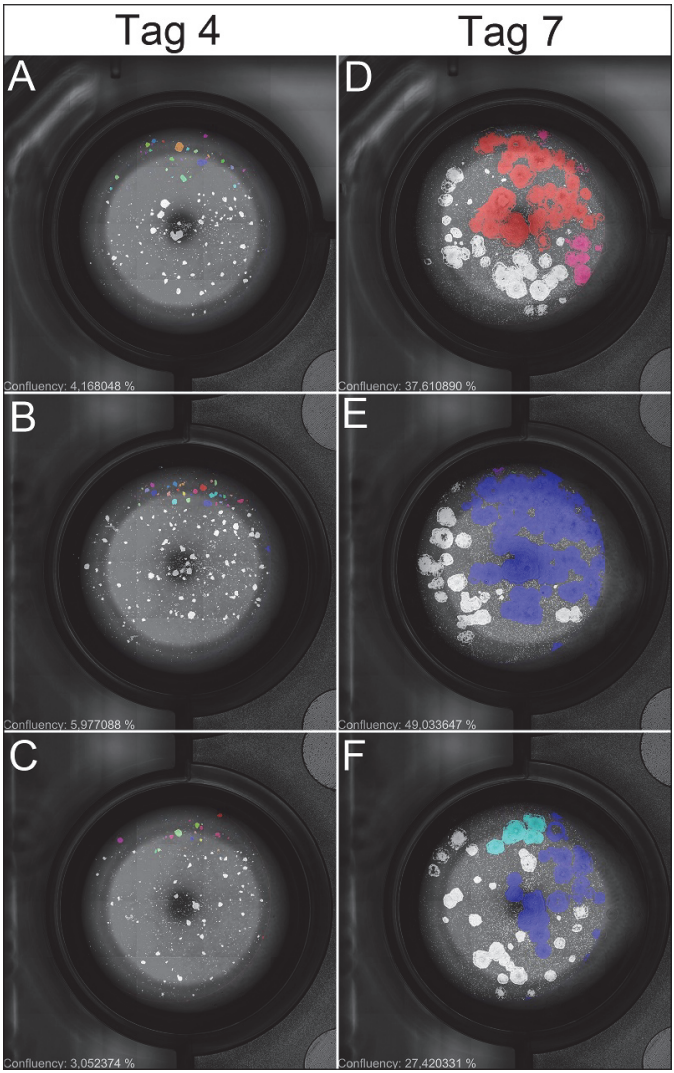


Abbildung 15.6: Ergebnisbilder der Konfluenzquantifizierung einer Phasenkontrastaufnahme in drei verschiedenen Wells einer 24-Well Mikrotiterplatte: (A-C) Aufnahme mit Ergebnis als Overlay an Tag 4 der Zellkultur, (D-F) Aufnahme mit Ergebnis als Overlay an Tag 7 der Zellkultur.

3 Zusammenfassung

Die bildverarbeitungsbasierte Konfluenzbestimmung von iPS-Zellkolonien nach dem vorgestellten Algorithmus ist ein Schlüssel für die effiziente Zellkultivierung vor allem in vollautomatisierten Bioproduktionsanlagen. Auf Basis der automatisiert ermittelten Konfluenzwerte kann die Prozesssteuerung eine rechtzeitige Passagierung bei Überschreitung eines gewissen Konfluenzwertes veranlassen, was eine Voraussetzung für ideale Wachstumsbedingungen der Zellkultur ist.

Literatur

1. U. Marx, F. Schenk, J. Behrens, U. Meyr, P. Wanek, W. Zang, R. Schmitt, O. Brüstle, M. Zenke und F. Klocke, „Automatic Production of Induced Pluripotent Stem Cells“, *Procedia CIRP*, Vol. 5, S. 2 – 6, 2013, First CIRP Conference on BioManufacturing.
2. K. Takahashi und S. Yamanaka, „Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors“, *Cell*, Vol. 126, Nr. 4, S. 663–676, 2006.
3. M. Kulik, F. Schenk und R. Schmitt, „Induziert pluripotente Stammzellen iPS – die vollautomatisierte Herstellung“, *GIT Labor-Fachzeitschrift*, Vol. 58, Nr. 2, S. 22–24, 2014.
4. S. Schmitz, *Der Experimentator: Zellkultur*, 3. Aufl. Heidelberg: Spektrum Akademischer Verlag, 2011.
5. G. Gstraunthaler und T. Lindl, *Zell- und Gewebekultur: Allgemeine Grundlagen und spezielle Anwendungen*, 7. Aufl. Berlin: Springer Spektrum, 2013.
6. F. Schenk, „High-Speed Microscopy – Scanning of Microtiter Plates at Unprecedented Speed“, *Imaging & Microscopy*, Vol. 16, Nr. 1, S. 20–22, 2014.
7. B. Jähne, *Digitale Bildverarbeitung*, 7. Aufl. Berlin: Springer, 2012.

Automatisierte Beurteilung der Schädigungssituation bei Patienten mit altersbedingter Makuladegeneration (AMD)

Stefan Kahl, Marc Ritter und Paul Rosenthal

Technische Universität Chemnitz,
Fakultät für Informatik,
Straße der Nationen 62, D-09111 Chemnitz

Zusammenfassung Die Analyse des Verlaufs des *Retinalen Pigmentepithels* (RPE) in OCT-Scans ist hilfreich, um die Schädigungssituation des Auges bei altersbedingter Makuladegeneration zu beurteilen. Zur Detektion des RPE-Verlaufs wird ein zweidimensionaler Bildverarbeitungsalgorithmus kreiert, der auf klassischen Operatoren beruht. Im weiteren Fokus dieses Beitrags steht die Entwicklung eines Tools, das die Annotation von RPE-Verläufen erlaubt, die einzelnen Stellgrößen des Detektionsalgorithmus automatisiert an einen vorhanden Datensatz anpasst und schließlich eine visualisierte Darstellung der Schädigungssituation ermöglicht.

1 Einleitung

Eine der am weitesten verbreiteten Augenerkrankungen ist die Schädigung der Netzhaut des menschlichen Auges im Bereich des schärfsten Sehens, der Makula. Dieses oft im Alter auftretende Krankheitsbild ist die häufigste Ursache für Erblindung bei Menschen über 50 Jahren. Der medizinische Kontext der Augenerkrankungen zeichnet sich vor allem durch hochspezialisierte Verfahren der Bildgebung aus. In den letzten Jahren hat sich bei der Früherkennung der altersbedingten Makuladegeneration und der Untersuchung der Ausprägung der Erkrankung die optische Kohärenztomographie (engl. *Optical Coherence Tomography*, kurz OCT) etabliert. Eine elementare Voraussetzung für die Bewertung der Schädigungssituation bei altersabhängiger Makuladegene-

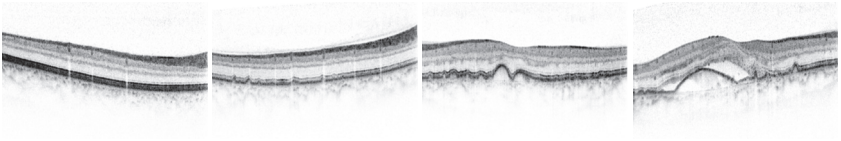


Abbildung 16.1: OCT-Scans mit gesundem Verlauf, leichter Schädigung, teilweise stark geschädigtem RPE und massiver Schädigung mit starken Deformationen

ration ist die Identifikation einzelner Schichten der Netzhaut in OCT-Aufnahmen. Bei der Einordnung einer Schädigungssituation spielen vor allem der Kantenverlauf, aber auch die flächenmäßige Ausbreitung und Textur der retinalen Schichten eine Rolle.

Anhand statistischer Modelle und Trainingsdatensets bestimmen *Kajić et al.* [1] den Verlauf solcher Schichten. Bei *Mayer et al.* [2] werden zunächst die inneren und äußeren Grenzen der Retina lokalisiert. Eine Identifikation der Netzhautschichten erfolgt anschließend über eine heuristische Auswertung des Intensitätsgefälles der Schichtgrenzen. *Fernández et al.* [3] nutzen weiterentwickelte Ansätze der Diffusion zur Rauschminderung und Kantenverstärkung für die Gewebeanalyse. Die Schichten der Netzhaut werden ebenfalls lokal, nach vorangehender Bestimmung der Netzhautgrenzen durch aufeinanderfolgende Peaks im Helligkeitsplot der Aufnahme bestimmt. Bei altersbedingter Makuladegeneration sind vor allem das an die stark durchblutete Aderhaut des äußeren Auges grenzende Retinale Pigmentepithel und die unmittelbar darüber liegenden inneren und äußeren Segmente der Sehzellen betroffen (vgl. *Klabe* [4, 143f]). Die zuverlässige Detektion des Verlaufs des RPE ist für die Bestimmung der Schädigungssituation bei Patienten mit AMD von entscheidender Bedeutung, wobei starke Deformationen sowie kleine, kontrastarme Beschädigungen dieser Gewebeschicht zu bewältigen sind. Bereits *Lee et al.* [5] untersuchen die Schädigung der RPE-Schichten bei 46 Augen von 33 Patienten in drei verschiedenen Kategorien mit Hilfe proprietärer Software der OCT-Gerätehersteller *Cirrus HD-OCT RPE Elevation Analysis* und *Carl Zeiss Meditec*. Hingegen liegt die Entwicklung eines zweidimensionalen Bildverarbeitungsalgorithmus mit klassischen Operatoren zur Detektion des RPE-Verlaufs und einer Klassifikation der Schädigungssituation im Fokus der vorliegenden Arbeit, wobei alle verfügbaren Stellgrößen automatisiert op-

timiert werden sollen.

Bei der optischen Kohärenztomographie wird die Netzhaut mit Licht von kurzer Kohärenzlänge, das von einer superlumineszenten Laser-Diode emittiert und über einen Strahlteiler geteilt wird, punktweise abgetastet (vgl. *Abouzeid & Wolfensberger* [6, S.1f]). Ein Teil des Lichts trifft auf einen Referenzspiegel, ein anderer auf die Netzhaut des untersuchten Auges. Das Licht wird vom Gewebe reflektiert und der Grad der Reflexion gibt Auskunft über die Beschaffenheit. Auf diese Weise ist es möglich, die Netzhaut nicht nur oberflächlich nach Schädigungen zu untersuchen. Auch tiefere Gewebeschichten können durch die Tomographie sichtbar gemacht werden. Diese Art der Bildgebung hat allerdings zur Folge, dass stark reflexive Strukturen im oberen (dem Innenauge zugewandten) Bereich ein Eindringen des Lichtstrahls in tieferliegende Regionen verhindern können. Aus diesem Grund entstehen vor allem durch Blutgefäße helle, vertikal verlaufende Streifen, die das RPE löchrig erscheinen lassen. Zusätzlich verursacht das sequentielle *Abtasten* der Netzhaut mit einem Lichtstrahl ein starkes Bildrauschen. Beide Eigenheiten der OCT-Scans sind aus Sicht der Bildverarbeitung nicht ideal. Das vorliegende Bildmaterial zweier gesunder und zweier kranker Augen mit je 19 OCT-Aufnahmen und einer Auflösung von 512×496 Pixel entstammt mit freundlicher Genehmigung dem *Universitätsklinikum Freiburg*.

2 Verfahren

Zielstellung bei der Entwicklung eines Verfahrens zur automatisierten Detektion des RPE war die Kombination einfacher schwellwertbasierter Methoden der Bildverarbeitung mit dem Kontext von OCT-Scans bei makulärer Degeneration. Das hier vorgestellte Verfahren entstand unter der Annahme, dass domänenspezifisches Wissen solche einfachen Methoden signifikant verbessern kann.

2.1 Detektion des RPE

Eine geeignete Vorverarbeitung von medizinischem Bildmaterial ist ausschlaggebend für den Erfolg der semantischen Merkmalsextraktion. Gerade bei OCT-Aufnahmen spielen Methoden zur Rauschentfernung,

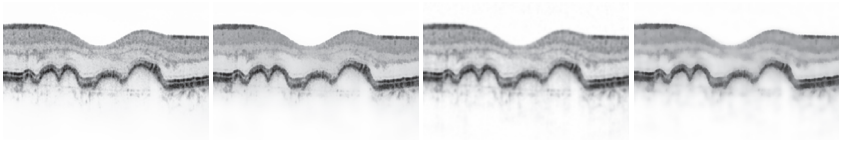


Abbildung 16.2: Bildausschnitt der ersten vier Iterationsstufen nichtlinearer Gaußscher Filterketten bei stark geschädigtem RPE

die kantenerhaltend arbeiten, eine große Rolle, wobei im vorgestellten Verfahren nichtlineare Gaußsche Filterketten zum Einsatz gelangen, wie sie von *Weule* [7] vorgestellt wurden (vgl. Abbildung 16.2). Eine weitere, wichtige Komponente der Vorverarbeitung ist die Spreizung des Histogramms der Helligkeitswerte. Die stark reflektierenden Strukturen des RPE lassen sich durch Verschiebung des Schwarz- und Weißpunktes im Histogramm noch deutlicher hervorheben, was eine zuverlässige Detektion dieser Partien mit Hilfe eines Schwellwertes ermöglicht.

Eine automatisierte Bewertung des Grades der Schädigung einer Netzhaut auf Basis von OCT-Aufnahmen setzt die zuverlässige Detektion des Verlaufs des RPE voraus. Als visuell diskriminierendes Merkmal wurde die deutlich sichtbare und oft kontrastreiche Kante zwischen RPE und Aderhaut selektiert, die es im folgenden zu detektieren gilt. Ein Großteil der vom RPE sichtbaren Bereiche lässt sich durch eine geeignete Vorverarbeitung bereits als zusammenhängende Segmente hervorheben. Zur Detektion dieser Bereiche ist ein einfacher Schwellwert ausreichend. Allerdings weißen vor allem stärker geschädigte Teilabschnitte ein geringeres Kontrastverhältnis auf und sind somit nicht trivial von anderen Gewebeschichten zu unterscheiden. Um trotzdem einen zusammenhängenden Kantenverlauf zu erhalten, der zudem möglichst genau den teilweise erheblichen Deformationen entspricht, wurden als Bewertungskriterien folgende Bedingungen für zum RPE gehörende Bildpunkte erstellt, die in Abbildung 16.3 beispielhaft illustriert sind:

1. Helligkeitswert des Pixels muss über dem Schwellwert liegen.
2. Y-Koordinate darf sich nicht mehr als wenige Punkte vom Vorgängerwert unterscheiden. (Plausibilitätsbedingung für einen zusammenhängenden Kantenverlauf)
3. In der Nachbarschaft müssen sich genügend Pixel befinden, die ebenfalls Bedingung 1 erfüllen. (Plausibilitätsbedingung für ein Segment einer gewissen Dicke)

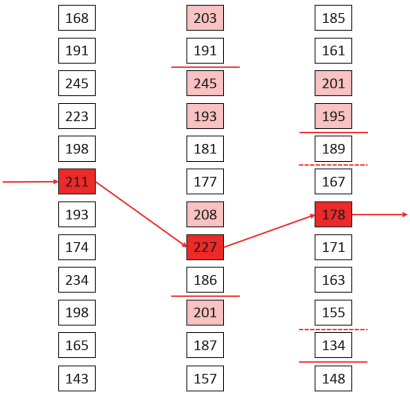


Abbildung 16.3: Verdeutlichung des Auswahlverfahren bei einem Schwellwert der Helligkeit von 192. In Spalte eins (links) bestimmt die Position des Ausgewählten Bildpunktes (211) die Region in Spalte zwei (Mitte, rote Linien), in der sich valide Pixel befinden können (Bedingung 2). Aus diesen Bildpunkten kommen vier Pixel in Frage (227, 208, 193, 245), da deren Helligkeitswert über dem Schwellwert liegt (Bedingung 1). Nur die Punkte mit den Helligkeitswerten 227 und 245 erfüllen jedoch Bedingung 3, der vertikal am nächsten zum Vorgänger platzierte Punkt wird laut Bedingung 4 gewählt (227). In Spalte drei (rechts) befindet sich kein Bildpunkt, der Bedingung 1-3 erfüllt. Daher wird laut Bedingung 4 (gestrichelte Linien) der Bildpunkt, welcher dem Schwellwert am nächsten ist (178) gewählt.

4. Werden multiple Punkte gefunden, die Bedingung 1-3 erfüllen so wird derjenige gewählt, dessen Y-Koordinate der des Vorgängers am nächsten ist. (Bedingung für einen möglichst homogenen Kantenverlauf)
5. Wird kein Bildpunkt gefunden, der Bedingung 1-3 erfüllt, so wird aus allen Punkten, die in Y-Richtung nicht weiter als wenige Pixel vom Vorgänger entfernt sind, derjenige gewählt, der dem Schwellwert am nächsten ist. (Bedingung zur Identifikation kontrastarmer Bereiche als Teilstück des RPE)

Die Bedingungen 2, 4 und 5 sind auf Informationen zum vorangegangenen, zum RPE gehörenden Bildpunkt angewiesen. Eine ungünstige Auswahl des initialen Startpunktes kann zu einer Fehldetektion im

Kantenverlauf führen. Um diesem entgegenzuwirken, wird zusätzlich die 5×5-Nachbarschaft eines jeden validen Startpunktes untersucht. Darüber hinaus können auch kontrastreiche Gewebeteile der an das RPE angrenzenden Aderhaut trotz Vorverarbeitung und Auswahlheuristik den detektierten Kantenverlauf stören. Mit Hilfe einer Verlaufskorrektur (Smoothing) lässt sich durch das Eliminieren vereinzelter Ausreißer nach *McMaster & Shae* [8] ein homogeneres Kantenbild erzielen, wobei auch kleine Beschädigungen des RPE erkennbar bleiben sollen.

2.2 Parameteroptimierung

Das Verfahren zur Detektion des Verlaufs des RPE profitiert sehr stark von einer geschickt gewählten Parametrisierung. Erfolg und Misserfolg werden in großem Maße durch die Schwellwerte zur Auswahl gültiger RPE-Pixel beeinflusst. Im Vordergrund der Untersuchung stand die Entwicklung eines geeigneten Tools zur Ermittlung optimaler Parameterkonstellationen. Einerseits erlaubt es dem Benutzer OCT-Aufnahmen zu annotieren, andererseits auf Basis dieser Metadaten automatisiert eine Optimierung der Einstellparameter durchzuführen und die Ergebnisse anschließend zu visualisieren (vgl. Abbildung 16.4).

Die Annotation der OCT-Aufnahmen erfolgt manuell durch einen geschulten Benutzer. Nicht in jedem Fall ist die Fachkompetenz des Betrachters ausreichend, um zweifelsfrei den Verlauf des RPE zu bestimmen. Herausforderungen treten vor allem in Bereichen auf, in denen sich das RPE visuell wenig von den umgebenden Gewebeschichten unterscheidet. Die Korrektheit der Annotation ist aber Voraussetzung für die Wahl geeigneter Parameter. In unserem Fall wurden die Bilder unter Hilfestellung einer Augenärztin durch die Autoren annotiert. Zur Unterscheidung von wenig, mäßig und stark geschädigten Bereichen erscheint diese Art der manuellen Analyse zunächst ausreichend.

Die Möglichkeit einer automatisierten, systematischen Optimierung des Parametersets ist eine Grundvoraussetzung für die Minimierung des Verfahrensfehlers. Die Anzahl möglicher Kombinationen aller beteiligten Variablen erscheint für die computergestützte Optimierung mittels erschöpfender Suche sehr groß. Um diesen Suchraum einzuschränken, wurden für alle Parameter empirisch ermittelte Intervalle definiert, in denen sich mit hoher Wahrscheinlichkeit optimale Werte

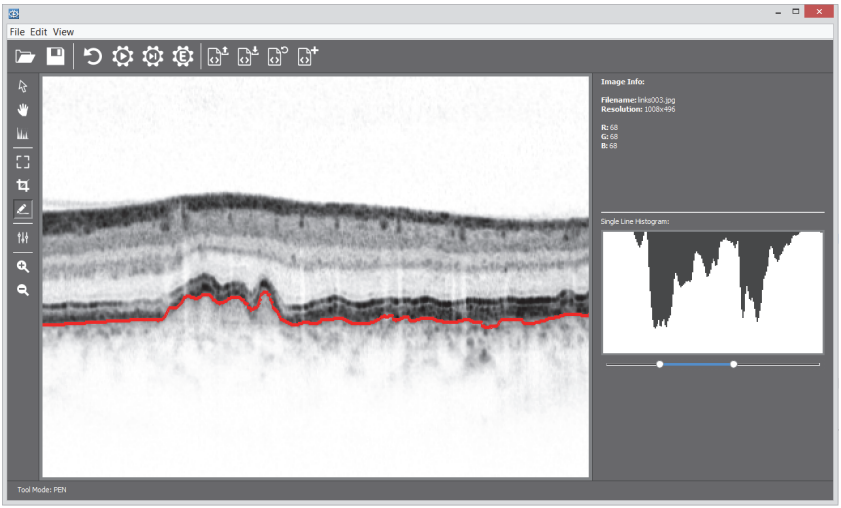


Abbildung 16.4: Graphische Benutzungsoberfläche des entwickelten Tools, mit dem OCT-Aufnahmen manuell annotiert, Metadaten als XML exportiert und automatisierte RPE-Detektionen (rote Linie) evaluierbar sind sowie die Stapelverarbeitung zahlreicher OCT-Scans ermöglicht wird.

für jede Variable befinden. Die Bestimmung des optimalen Parametersets für jedes Einzelbild erfolgt durch die Bestimmung der Abweichung des in jedem Iterationsschritt detektierten RPE-Verlaufs vom manuell annotierten Verlauf. Der kumulative Fehler ergibt sich aus der Differenz der Y-Koordinaten im paarweisen Vergleich aller Pixel entlang der X-Achse: Umso geringer die Abweichung der automatischen Detektion von der manuellen Annotation ausfällt, desto geringer ist der Fehler. Das Parameterset, das im direkten Vergleich beider RPE-Verläufe den geringsten Fehler liefert, wird als optimale Wahl betrachtet.

Die automatisierte Optimierung der Variablen des Verfahrens hat im Wesentlichen drei Kernparameter offenbart. Der Vorteil der Vorverarbeitung der Aufnahmen durch nichtlineare Gaußsche Filterketten ist bei einer Iterationstiefe von vier am größten. Die Verringerung des Rauschens bei gleichzeitiger Wahrung der Konturen ist Grundvoraussetzung für eine Detektion mit geringem Fehler. Durch die Spreizung des Histogramms der Helligkeitswerte und der Anpassung des zu-

gehörigen Schwellwertes werden zusammenhängende Bereiche sichtbar und kleinere Lücken bereits in der Vorverarbeitungsstufe geschlossen. Allerdings verstärkt diese Art der Vorverarbeitung auch einzelne Gewebepartien, die in der Aderhaut unterhalb des RPE zu finden sind und unter Umständen zu Fehlern bei der Bestimmung des RPE-Verlaufs führen können. Durch das Zusammenspiel von Histogrammspreizung und Klassifikation einzelner Pixel durch einen Schwellwert können im vorliegenden Ansatz gute Ergebnisse erzielt werden.

2.3 Beurteilung der Schädigung

Eine Kategorisierung der als RPE identifizierten Teile einer OCT-Aufnahme erfolgt anhand des Kurvenverlaufs der als zusammenhängende Markierung erfassten Einzelpixel. Bei der Bewertung spielen globale Merkmale, die den Verlauf in seiner Gesamtheit betrachten, als auch lokale Eigenschaften, die wiederum nur einen Teilabschnitt beschreiben, eine Rolle. Es hat sich gezeigt, dass vor allem letztere als gutes Maß zur Bewertung der Schädigung des RPE geeignet sind.

Dabei wird der detektierte Verlauf in Teilbereiche zerlegt und jeder einzelne Abschnitt auf Unregelmäßigkeiten untersucht, wobei die Differenz zwischen maximaler und minimaler Y-Position der einzelnen Pixel des Abschnitts und somit die Abweichung vom geraden bis leicht gekrümmten *gesunden* Verlauf als primäres Maß für die Bewertung Verwendung findet. Eine Ausdehnung dieser lokalen Merkmale auf die gesamte Detektion ist allerdings nicht in jedem Fall sinnvoll, da oftmals auch Aufnahmen von gesunden Partien der Netzhaut erhebliche Krümmungen aufweisen. Eine Unterscheidung auf Basis von Min-Max-Vergleichen würde zu Fehleinschätzungen führen, weshalb der alleinige Einbezug von lokalen Eigenschaften ungeeignet erscheint, da einzelne Fehldetektionen in Form von *Ausreißern* das Schädigungsbild verfälschen können. Hingegen lässt die Betrachtung der Gesamtheit aller lokalen Unebenheiten Rückschlüsse auf das Schädigungsbild zu. Übersteigt die Zahl der Abweichungen einen kritischen Wert im gesamten Verlauf, liegt die Vermutung nahe, dass es sich um beschädigtes Gewebe handelt.

Auch bei der Klassifikation finden geeignete Schwellwerte Einsatz, indem diese so vorselektiert werden, dass der vorhandene Datenbestand möglichst zuverlässig erkannt wird. Demnach ist davon auszu-

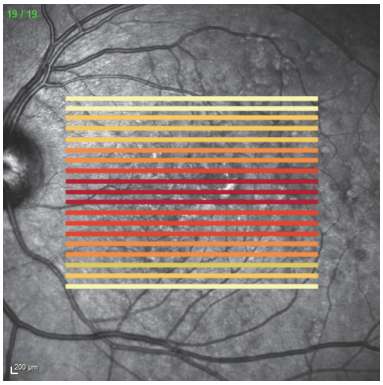


Abbildung 16.5: Visualisierung der detektierten Schädigungen des RPEs in den einzelnen Schichten von hell (keine) bis dunkel (massive Schädigung).

gehen, dass das Verfahren auch auf unbekanntem Bildmaterial gleichen Typs zuverlässig arbeitet. Problematisch erscheint eine Abweichung der Scans aus technischer Sichtweise, so zum Beispiel bei anderen Gerätetypen mit stark veränderter Auflösung oder stärkerem Kontrastverhältnis. Darüber hinaus könnten auch komplexe Krankheitsbilder nicht immer korrekt erkannt werden. Besonders hervorzuheben ist hier der lokale Durchriss des RPE bei weit fortgeschrittenem Krankheitsbild. Weitere Schwierigkeiten treten ebenfalls bei der Abgrenzung zwischen nicht geschädigten und nur leicht geschädigten Bereichen aufgrund der Verallgemeinerungen der Einstellgrößen und Schwellwerte auf. Jedoch lassen sich mäßig bis stark geschädigte Bereiche sehr zuverlässig unterscheiden. Bei der Wahl der Visualisierung stand vor allem eine leichte Erkennbarkeit der Schädigungssituation des gesamten Auges und nicht nur der eines einzelnen Scans im Vordergrund (siehe Abbildung 16.5). Die Art der Visualisierung ist prototypisch, verdeutlicht aber bereits zum aktuellen Zeitpunkt das Analyseergebnis. Zukünftig sollen Mediziner auf einen kurzen Blick die Schwere der Erkrankung feststellen können. Da auch hier die Gesamtheit des Ergebnisses wichtiger ist als die individuelle Korrektheit einzelner Scans, erscheint eine minimale Abweichung der Detektion von den tatsächlichen RPE-Verläufen unkritisch, solange der Mediziner anhand der automatisierten Auswertung entscheiden kann, ob Handlungsbedarf besteht.

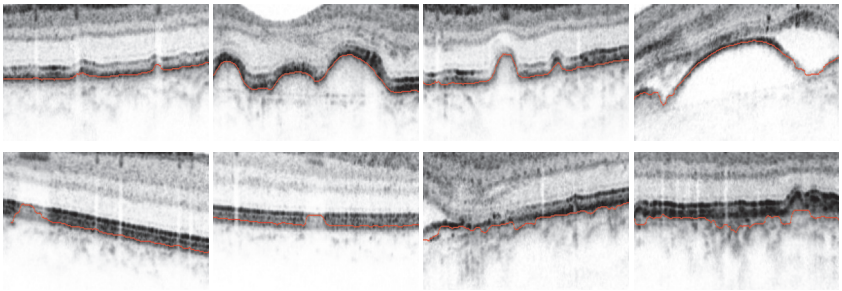


Abbildung 16.6: Korrekte Detektionen (oben): geringfügige Deformation, starke Schädigung mit hohem Kontrast, mäßige Schädigung bei geringem Kontrast, massive Deformation bei geringem Kontrast an den Rändern. Fehlerhafte Detektionen (unten): geringer Kontrast aufgrund eines Blutgefäßes in einem gesunden Auge; Unterbrechung durch ein Blutgefäß; kontrastreiche Gewebestrukturen unterhalb des RPE; Aderhaut und RPE lassen sich visuell nicht deutlich genug voneinander abgrenzen.

3 Evaluation

Das automatische Ausprobieren aller möglichen Parameterkombinationen liefert für jedes Bild ein optimales Parameterset, mit dem der Verlauf des RPE und der Grad der Schädigung sehr genau bestimmt werden kann (vgl. Tabelle 16.1). Leider lassen sich diese optimalen Sets von Parametern selten verallgemeinern. Unter Berücksichtigung der händischen Annotation, wurde jenes Parameterset als globale Einstellung gewählt, das in der Gesamtheit aller Bilder mit RPE-Schädigung den geringsten kumulativen Fehler aufweist. Dabei muss die Wahl geeigneter Parameter ebenso berücksichtigen, dass der Fehler bei einer einzelnen Aufnahme sehr groß sein kann, auch wenn die Summe aller Fehler minimal erscheint, wodurch im Einzelfall das RPE mit kleinen fehlerhaften Passagen erkannt, jedoch der Grad der Schädigung nur unzureichend bestimmt wird. Dieser Umstand fällt vor allem bei der Unterscheidung von nicht geschädigten Bereichen und nur leicht geschädigten Strukturen ins Gewicht. Insbesondere bei gesunden Augen steigt die Fehlerrate aufgrund von Unregelmäßigkeiten im Bildmaterial (siehe Abbildung 16.6). Da der Fokus aber auf der Detektion geschädigter Bereiche liegt und der Gesamteindruck der Schädigungs-

Testbild	OCT-Layer	∅ Fehler individuell bestes Set	∅ Fehler allgemein bestes Set	∅ Fehler angewandtes Set
Gesund I	04	0,5313	0,5840	0,6680
Gesund II	16	0,6309	2,3633	2,5078
Krank links I	05	0,9531	1,3301	1,0840
Krank links II	10	0,8926	1,2070	1,1699
Krank rechts I	04	1,5801	1,8750	1,7305
Krank rechts II	11	1,5469	1,9512	1,8184
↑ 6 Testbilder	–	1,0225	1,5518	1,4964

Tabelle 16.1: Durchschnittliche Abweichung des detektierten RPE-Verlaufs vom annotierten Groundtruth auf einer Auswahl an Bildern bestimmter Kategorien. Es ist zu erkennen, dass das individuelle Set sehr gute Ergebnisse mit durchschnittlich etwas mehr als einem Pixel Abweichung liefert, wovon die allgemeingültige beste Parametrierung über die Gesamtmenge aller OCT-Scans stärker abweicht. Das angewandte Parameterset (vgl. Abbildung 16.4) schneidet bei gesunden Augen schlechter, bei geschädigten dafür besser ab.

situation wesentlich entscheidender ist als einzelne schadhafte Strukturen, wurde bei der Parametrisierung die Priorität auf eine Minimierung des Fehlers bei mäßig bis extrem stark geschädigten Strukturen gelegt. Unter der Annahme, dass gesunde Augen nur in geringer Zahl untersucht werden und die Korrektheit der Bestimmung der Schädigungssituation vor allem für die Kontrolle des Therapieverlaufs wichtig ist, erscheint dieses Vorgehen plausibel. In Zukunft ist es denkbar, unterschiedliche Parametersets auf dem Bildmaterial anzuwenden und für die Analyse jene Heuristik auszuwählen, die mit der größten Wahrscheinlichkeit den Verlauf des RPE korrekt abbildet.

4 Zusammenfassung und Ausblick

Durch die geschickte Kombination klassischer Bildverarbeitungsoperatoren ist es möglich, den Verlauf des RPE in OCT-Aufnahmen zu rekonstruieren, wobei das kontrastreiche Erscheinungsbild eine zuverlässige Identifikation erlaubt. Eine geeignete Parametrierung verbessert das Ergebnis auch in für den Menschen schlecht zu erkennenden Berei-

chen. Plausibilitätskriterien für die Auswahl der zum RPE gehörenden Bildpunkte erweitern die Low-level-Merkmale um eine semantische Komponente. Die Kombination mit globalen und lokalen Bewertungskriterien lässt eine computergestützte Aussage über den Grad der Schädigung bei Patienten mit altersbedingter Makuladegeneration zu. Derartige Gewebeuntersuchungen erlauben aber nicht immer eindeutige Rückschlüsse auf die Art der Beeinträchtigung des Sehvermögens. Die Anwendbarkeit der automatisierten Beurteilung und Darstellung des Schädigungsgrades muss durch eine Befragung von Patienten und Medizinern eruiert werden, wobei die Schädigung der Photorezeptoren eine große Rolle spielt. Letztlich ist eine Erweiterung des Verfahrens durch Einbezug von persönlichen Patientendaten (Vorerkrankungen, Risikofaktoren, Sehtest) und medizinischer Analysen denkbar.

Literatur

1. V. Kajíc, M. Esmaeelpour, C. Glittenberg, M. Kraus, J. Hornegger, R. Othara, S. Binder, J. G. Fujimoto und W. Drexler, „Automated three-dimensional choroidal vessel segmentation of 3D 1060 nm OCT retinal data“, *Biomedical Optics Express*, Vol. 4, Nr. 1, S. 134–150, 2012.
2. M. A. Mayer, J. Hornegger, C. Y. Mardin und R. P. Tornow, „Retinal nerve fiber layer segmentation on fd-oct scans of normal subjects and glaucoma patients“, *Biomed. Opt. Express*, Vol. 1, Nr. 5, S. 1358–1383, Dec 2010.
3. D. C. Fernández, H. M. Salinas und C. A. Puliafito, „Automated detection of retinal layer structures on optical coherence tomography images“, *Optics Express*, Vol. 13, Nr. 25, S. 10 200–10 216.
4. K. Klabe, „Pathophysiologie und Klinik der Makuladegeneration: Angiogenesehemmer als neue Therapieoption“, *Fortbildungstelegramm Pharmazie*, Vol. 1, S. 141–149, 2007.
5. S. Y. Lee, P. F. Stetson, H. Ruiz-Garcia, F. M. Heussen und S. R. Sadda, „Automated characterization of pigment epithelial detachment by optical coherence tomography“, *Invest Ophthalmol Vis Sci.*, Vol. 52, Nr. 1, S. 164.
6. H. Abouzeid und T. J. Wolfensberger, „Optical Coherence Tomography Assessment of Macular Oedema“, S. 1–18.
7. J. Weule Sievert, „Iteration nichtlinearer Gauß-Filter in der Bildverarbeitung“, Dissertation, Düsseldorf, 1994, 117 S.
8. R. B. McMaster und K. S. Shea, *Generalization in Digital Cartography*. Washington, D.C.: Assoc. of American Geographers, 1992.

Superpixel-gestützte Klassifikation von Stechmückengattungen mit der Bags-of-Features-Methode

Paul Grigoriev, Jonas Jäger, Christoph Kornek, Viviane Wolff und Klaus Fricke-Neuderth

Hochschule Fulda, Fachbereich Elektrotechnik und Informationstechnik, Marquardstr. 35, D-36039 Fulda

Zusammenfassung Gegenstand dieser Arbeit ist die superpixelgestützte Klassifikation unterschiedlicher Stechmückengattungen. Die Superpixel nehmen die Rolle der Merkmale in der Bag-of-Features (BoF) Methode ein. Für die Berechnung der Superpixel wird eine modifizierte Form des SLIC-Algorithmus verwendet. Die durchgeführten Modifikationen werden vorgestellt. Ferner werden drei einfache Varianten zur Berechnung der Deskriptoren aus den Superpixeln angeboten. Alle drei Varianten werden auf ihre Zuverlässigkeit bei der Klassifikation von drei Stechmückengattungen untersucht und mit den etablierten SIFT-Merkmalen verglichen. An einer Anzahl von Stichproben, bestehend aus den drei Gattungen Aedes, Anopheles und Culex, wird gezeigt, dass die Kombination von Superpixeln mit der BoF-Methode Trefferquoten von bis zu 99 % liefern kann.

1 Einleitung

In Zeiten globaler Erderwärmung steigt die Wahrscheinlichkeit dafür an, dass Krankheitserreger, die durch bestimmte Stechmücken hauptsächlich in tropischen Regionen übertragen werden, sich auch in Regionen wie Europa und damit dorthin ausbreiten, wo diese bisher aufgrund des Klimas nicht überleben konnten. Eine mögliche Folge wären Infektionskrankheiten wie Malaria, Dengue-Fieber und Gelbfieber. [1]

Um Infektionsrisiken im Vorfeld zu erkennen, müssen Überwachungssysteme zur Kontrolle der Population von Stechmückengattun-

gen (z. B. Aedes oder Anopheles), die zur Übertragung solcher Krankheiten in Frage kommen, entwickelt und etabliert werden. Ein solches System könnte eine Stechmückenfalle in Verbindung mit einer hochauflösenden CCD-Kamera und einem eingebetteten System zur automatischen Identifizierung der Stechmückengattung darstellen.

Die automatische Identifizierung von Stechmücken gestaltet sich komplex. Die Stechmücken haben keine feste Größe und Farbe. Die Positionen der Beine und Flügel im Bild können je nach Lage der Stechmücke stark variieren, darüber hinaus kann der Blickwinkel auf die Stechmücke bei der Aufnahme innerhalb einer Stechmückenfalle nicht explizit vorgegeben werden. Hinzu kommt, dass die einzelnen Gattungen sich nur unwesentlich voneinander unterscheiden.

Bei der manuellen Identifizierung von Stechmücken werden einzelne Körperteile mit Hilfe eines Identifikationsschlüssels auf bestimmte Kriterien hin untersucht und daraus die Gattung bestimmt. Diese Vorgehensweise wurde bei der Wahl der Algorithmen zur automatischen Identifizierung berücksichtigt, woraus die Fokussierung auf das Superpixel-Verfahren resultiert. Die Superpixel erlauben eine konturgenaue Erkennung der Lage einzelner Körperteile im Bild, die mit den üblichen Merkmalsextraktionsverfahren wie SIFT [2] nicht möglich ist, weil diese sich nur auf herausragende Merkmale im Bild beschränken. Diese Arbeit betrachtet daher, inwiefern sich Superpixel als Merkmale in der Bof-Methode für die Klassifikation von Stechmücken der drei Gattungen Aedes, Anopheles und Culex eignen. Die Erkenntnisse aus dieser Arbeit sollen dazu dienen, die Superpixel-gestützte Klassifikation auf einzelne Körperteile der Stechmücken auszuweiten.

2 Stand der Forschung

2.1 Superpixel-gestützte Klassifikation

Im Bereich der Klassifikation mit Hilfe von Superpixeln existieren einige aktuelle Arbeiten. In vielen Fällen werden, wie auch in dieser Arbeit, der k-means basierte SLIC-Algorithmus (Simple Linear Iterative Clustering [3]) für die Berechnung der Superpixel und eine Support-Vector-Machine [4] für die Klassifizierung verwendet.

In [5] werden geographische Veränderungen durch die automatische Auswertung von hochauflösenden Bildern erkannt. Hier werden

zunächst Superpixel gebildet und in einem weiteren Schritt als *verändert* oder *nicht-verändert* klassifiziert. Auch die Arbeiten [6–8] folgen dem Muster, Superpixel zu bilden und anschließend jeden Superpixel zu klassifizieren.

Im Gegensatz dazu wird in dieser Arbeit der Ansatz verfolgt, die aus den Superpixel berechneten Deskriptoren für die Klassifikation eines Objektes zu verwenden.

2.2 Klassifikation von Insekten

Die Möglichkeiten der modernen Bildverarbeitungsalgorithmen zur Identifikation von Insekten werden durch die jüngsten Arbeiten verdeutlicht.

In [9] wurde ein System zur automatischen Identifikation von Steinfliegenlarven vorgestellt. Die Identifikation der Steinfliegenlarven bestand aus der Detektion von „Regions of interest", der Beschreibung dieser Regionen mit den SIFT-Deskriptoren, der Erstellung eines Wörterbuchs aus den SIFT-Deskriptoren und der anschließenden Klassifikation mit „Logistic model trees". Die Arbeit folgte im Wesentlichen der Bag-of-Features-Methode und erreichte Erkennungsraten von bis zu 95 %.

In [10, 11] wurde die Arbeit aus [9] fortgesetzt. Um die Fehleranfälligkeit der Klassifikation zu senken, wurde in [10] auf die Erzeugung des Wörterbuchs verzichtet. Dieser Schritt wurde durch das Klassifikationsverfahren „Random forest trees" ersetzt. Der Schwerpunkt in [11] lag auf der Entwicklung eines Verfahrens, welches die Effizienz der Identifikation von Steinfliegenlarven verbessert und dennoch eine hohe Genauigkeit liefert. Das entwickelte Verfahren basiert auf der Extraktion von Merkmalen mit der „Haar random forests"-Methode kombiniert mit Support-Vector-Machine. Die durchschnittliche Zeit von der Ausführung bis zur Erzeugung des Histogramms konnte in Vergleich zu [10] um zwei Größenordnungen auf 5.03 Sekunden pro Bild bei einer ähnlichen Genauigkeit reduziert werden.

Eine weitere Arbeit mit dem Ansatz der Merkmalsextraktion wurde in [12] vorgestellt. Für die Extraktion von Merkmalen aus Insektenflügeln zur Klassifizierung verschiedener Libellenfamilien wurde die „hybrid moment invariants"-Methode verwendet. Als Klassifizierer kam Support-Vector-Machine zum Einsatz. Dieses Verfahren erreich-

te bei einer geringen Anzahl an Stichproben Genauigkeiten zwischen 72.5 % und 100 %.

Auch in [13] wurde mit Merkmalen zur Identifikation verschiedener Insektenarten gearbeitet. Die Merkmale wurden u. a. aus der Rechteckigkeit, Dehnung, Rundung, Exzentrik und den sieben „Hu moment"-Invarianten ermittelt. Als Klassifikator wurde „Random Trees" eingesetzt. Auch dieses Verfahren erreichte bei einer geringen Anzahl an Stichproben Genauigkeiten zwischen 80% und 100 %.

Die hohen Erkennungsraten aus den vorangegangenen Arbeiten, deren Grundlage stets die Merkmalsextraktion war, waren Anlass dafür, die BoF-Methode mit Superpixeln für die Klassifikation der drei Stechmückengattungen Aedes, Anopheles und Culex zu kombinieren.

3 Superpixel-Implementierung

Die Segmentierung des Bildes wird mit Hilfe von Superpixeln durchgeführt. Hierzu wird eine modifizierte Form des in [3] vorgestellten SLIC-Algorithmus verwendet. Die durchgeführte Anpassung des Algorithmus beinhaltet die Veränderung der Distanzmessung sowie die Reduzierung der Abbruchbedingung zu einer festen Anzahl an Iterationen. Ein genaue Erklärung des Algorithmus ist in [3] zu finden.

3.1 Distanzmessung

In die Distanzmessung fließt sowohl die räumliche Entfernung von einem Pixel P_k zu einem Cluster-Zentrum $C_k = [r_k, g_k, b_k, x_k, y_k]^T$ als auch der Farbunterschied ein. Die Berechnung der Distanz D erfolgt auf Basis einer 5-dimensionalen Manhattendistanz. Diese wurde gewählt, da sie einfacher als eine euklidische Distanz zu berechnen ist und im Hinblick auf die Erkennungsraten gleiche Ergebnisse liefert. Als Farbraum wird der RGB-Farbraum verwendet.

$$d_c = |r_i - r_j| + |g_i - g_j| + |b_i - b_j| \tag{17.1}$$

$$d_s = |x_i - x_j| + |y_i - y_j| \tag{17.2}$$

$$D = \alpha * \frac{d_c}{N_c} + \frac{d_s}{N_s} \tag{17.3}$$

Wie in den obigen Gleichungen zu sehen ist, gibt d_c den Farbunterschied und d_s die Entfernung der zu vergleichenden Pixel an. Für die Normierung der Farbdistanz wird die maximal mögliche Abweichung gewählt. Bei einer Farbtiefe von 8-Bit im RGB-Farbraum entspricht diese $N_c = 255 * 3$. Die räumliche Distanz wird auf $N_s = 2S$ normiert. N_s entspricht somit der maximalen Entfernung eines Pixels P_k zu einem Zentrum C_k in einer $2S \times 2S$ Umgebung um C_k. Der Faktor α wurde eingeführt, um den Einfluss des Farbunterschieds auf die Clusterbildung kontrollieren zu können. Wird α groß gewählt, so orientieren sich die Superpixel stark an den Farbregionen im Bild. Bei kleinem α werden die Superpixel in Form und Größe gleichmäßiger.

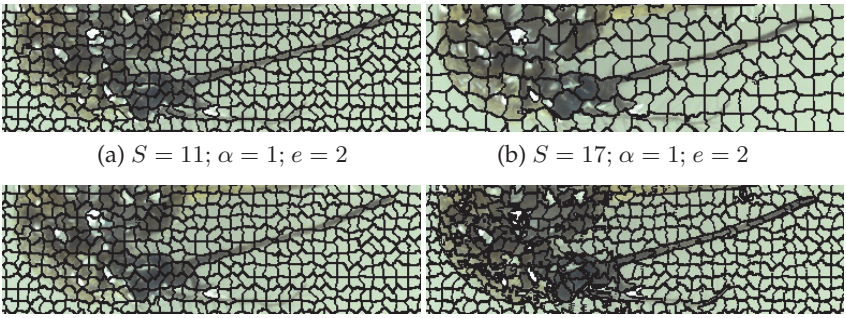
3.2 Parametrisierung bei Mückenbildern

Das Ziel der Segmentierung von Mückenbildern mit Superpixeln besteht darin, Pixel ähnlicher Farbe in einer bestimmten Umgebung zusammen zu fassen. Besonderes wichtig ist dabei, dass sich die Superpixel an die Anatomie der Mücke anpassen. Superpixelkonturen und Mückenkonturen sollen also an den Nahtstellen identisch sein. Um diese Forderung zu erfüllen, wurde eine Untersuchung der Segmentierungsergebnisse mit verschiedenen Algorithmus-Parametern durchgeführt. Hierzu sind die Parameter S und α sowie die Anzahl der Iterationen e systematisch angepasst und die Ergebnisse qualitativ verglichen worden.

Der Parameter S bestimmt maßgeblich die Größe der Superpixel. Die Variation von S ergab, dass die optimale Segmentierung bei einer Superpixelbreite, die der Breite eines Mückenbeins entspricht, gegeben ist (siehe Abb. 17.1(a)). Wird S kleiner gewählt, so vergrößert sich die Anzahl der Superpixel. Wird S größer gewählt, so ist die Anpassung der Cluster an den Mückenkörper nicht mehr optimal (Abb. 17.1(b)).

Mit Hilfe des Parameters α kann der Einfluss des Farbunterschieds auf die Clusterbildung gesteuert werden. Wie in Abbildung 17.1(c) zu erkennen ist, passen sich die Superpixel für kleine Werte von α nicht mehr korrekt an die Konturen der Mücke an. Man beachte die Cluster im Bereich des Rüssels der Mücke. Werden größere Werte für α gewählt, so gleichen sich die Clustergrenzen immer feineren Strukturen des Bildes an (siehe Abb. 17.1(d)).

Die Anzahl der Iterationen bestimmt darüber, wie oft eine Anpassung



(a) $S = 11$; $\alpha = 1$; $e = 2$ (b) $S = 17$; $\alpha = 1$; $e = 2$

(c) $S = 11$; $\alpha = 0.5$; $e = 2$ (d) $S = 11$; $\alpha = 8$; $e = 2$

Abbildung 17.1: Kopf und Thorax einer Stechmücke der Art Aedes aegypti. Segmentierung mit verschiedenen Werten des Parameters S. (a, b) Segmentierung mit verschiedenen Werten des Parameters α. (c, d)

der Superpixel durchgeführt wird. Eine ausreichende Anpassung der Superpixel an homogene Farbregionen der Mücke ist bereits nach zwei Iterationsschritten erreicht. Wird die Anzahl der Iterationen erhöht, so findet keine signifikante Verbesserung der Clusteranpassung statt.

4 Klassifikation mit Superpixeln

Um eine Klassifikation von Stechmücken durchzuführen, wurde die Bag-of-Features-Methode [14] gewählt.

Die Adaption der BoF-Methode wird in Abbildung 17.2 veranschaulicht. Zu jeder Klasse existieren Referenzbilder. Auf jedem Bild ist nur das Objekt zu sehen. Der Hintergrund wurde manuell zu weiß gesetzt. Die Referenzbilder werden in Trainingsbilder und Testbilder aufgeteilt. Beim Training werden aus allen Trainingsbildern die Superpixel extrahiert. Diese schließen nur die Teile des Bildes mit dem zu klassifizierenden Objekt ein, der weiße Hintergrund wird bei der Generierung der Superpixel ignoriert. Jedes Superpixel wird durch einen Deskriptor beschrieben. Um einen geeigneten Deskriptor zur Beschreibung eines Superpixels zu finden, wurden 3 unterschiedliche Varianten untersucht.

- Superpixelzentrum C_k als Grauwert, berechnet durch $Y = 0.3 \cdot R + 0.59 \cdot G + 0.11 \cdot B$

- Superpixelzentrum C_k als Grauwert und die Größe des Superpixels in Pixeln
- RGB-Farbwerte eines Superpixelzentrums C_k und die Größe des Superpixels in Pixeln

Auf der Basis der Deskriptoren wird das Codebook erzeugt. Dazu werden alle Deskriptoren aus den Trainingsbildern mit k-Means [15] geclustert. Das Codebook wird benötigt, um zu jedem Trainingsbild ein Histogramm zu erstellen, welches die Häufigkeit eines Clusters aus dem Codebook in einem Bild wiedergibt. Hierfür wird zu jedem Deskriptor aus einem Trainingsbild das Clusterzentrum mit der kürzesten Entfernung im Codebook gesucht. Anschließend werden die Histogramme mit einem Klassenlabel versehen und ein Klassifikator trainiert.

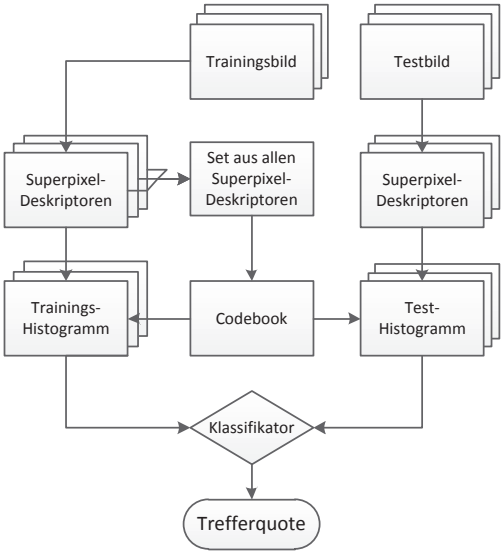


Abbildung 17.2: Adaption der BoF-Methode.

Um die Qualität des trainierten Klassifikators zu testen, werden die Testbilder herangezogen. Zu jedem Testbild werden die Deskriptoren aus den Superpixeln berechnet. Anschließend wird mit Hilfe des beim Training erzeugten Codebooks und den Deskriptoren ein Histogramm

erstellt. Dieses Histogramm wird mit dem Klassifikator ausgewertet. Ist die Trefferquote des Klassifikators über alle Testbilder zu gering, so wird in einem neuen Training die Anzahl der Trainingsbilder vergrößert sowie die Größe des Codebooks und die Lage der Cluster im Codebook variiert, bis die höchstmögliche Trefferquote erzielt wird.

5 Resultate

Um das vorgestellte Verfahren zu evaluieren, wurden Experimente mit verschiedenen Deskriptoren und Trainingsbildern durchgeführt.

Als Testmenge wurden insgesamt 103 Bilder von Stechmücken verwendet. Darunter befinden sich 37 Bilder der Gattung Aedes, 22 Bilder der Gattung Anopheles und 44 Bilder der Gattung Culex. Alle Bilder stammen von Insekten, die im Labor gezüchtet wurden. Die Auflösung der Bilder beträgt 800×600 Pixel. Der Hintergrund wurde auf jedem Bild manuell entfernt, da der Schattenwurf der Mücke auf den Hintergrund und der Farbverlauf des Hintergrundes die Ergebnisse verfälscht hätten. Position und Rotation der Mücken auf den Bildern sind zufällig.

Jeder Test, bestehend aus dem Training und der Klassifikation der Gattungen, wurde mit den drei vorher erwähnten Deskriptoren (Grauwert, Grauwert+Größe, Farbe+Größe) und den etablierten SIFT-Merkmalen durchgeführt. Zur Ermittlung der SIFT-Merkmale und der Berechnung von SIFT-Deskriptoren wurden die Standardfunktionen des OpenCV-Frameworks verwendet. Die Größe des Codebooks wurde auf 150 Wörter festgelegt. Diese Größe wurde empirisch ermittelt.

Support-Vector-Machine wurde als Klassifikator eingesetzt. Der Typ der SVM ist „C-Support Vector Classification“. Der Typ des Kernels ist die „Radial basis“-Funktion (RBF). Die Abbruchbedingung wurde auf 100 Iterationen bzw. auf eine Genauigkeit von $1e{-}6$ eingestellt. Der Parameter Gamma wurde in Abhängigkeit von der Codebookgröße berechnet: $1/(codebooksize \cdot 10)$. Die Variable C des Optimierungsproblems wurde auf 8 eingestellt.

Es wurden insgesamt 3 Testreihen durchgeführt. In der ersten Testreihe (Tabelle 17.1) wurde nur ein Trainingsbild pro Gattung verwendet, in der zweiten (Tabelle 17.2) zwei und in der dritten (Tabelle 17.3) wurden drei Trainingsbilder (Tabelle 17.3) verwendet. Die Lage der Mücke auf dem Bild war entscheidend für die Wahl der Trainingsbilder. Es wurde

darauf geachtet, dass Stechmücken der selben Gattung in verschiedenen Lagen gewählt wurden (z. B. eine Mücke in Rückenlage und eine andere Mücke der selben Gattung in Bauchlage). Alle Versuche wurden mit den gleichen Testbildern durchgeführt.

Deskriptor	Aedes	Anoph.	Culex	Erkennungsrate
Grauwert	24/37	7/22	43/44	64,8%
Grau.+Größe	22/37	13/22	41/44	70,57%
Farbe+Größe	33/37	12/22	39/44	77,45%
Sift	35/37	22/22	27/44	85,32%

Tabelle 17.1: Erkennungsraten für ein Trainingsbild pro Gattung.

Deskriptor	Aedes	Anoph.	Culex	Erkennungsrate
Grauwert	25/37	16/22	42/44	78,58%
Grau.+Größe	25/37	21/22	40/44	84,64%
Farbe+Größe	30/37	21/22	42/44	90,66%
Sift	34/37	22/22	31/44	87,45%

Tabelle 17.2: Erkennungsraten für zwei Trainingsbilder pro Gattung.

Deskriptor	Aedes	Anoph.	Culex	Erkennungsrate
Grauwert	26/37	19/22	42/44	84,03%
Grau.+Größe	29/37	20/22	41/44	87,49%
Farbe+Größe	37/37	22/22	43/44	99,24%
Sift	36/37	22/22	20/44	80,91%

Tabelle 17.3: Erkennungsraten für drei Trainingsbilder pro Gattung.

Die normierte Erkennungsrate eines Deskriptors wurde berechnet, indem die Erkennungsraten aller 3 Gattungen aufaddiert und anschließend durch 3 geteilt wurden. Wie die vorgestellten Tabellenwerte zeigen, verbessert sich die Erkennung mittels Superpixel-Deskriptoren für eine steigende Anzahl von Trainingsbildern. Im Fall der Klassifizierung auf Basis der SIFT-Merkmale kann keine Verbesserung festgestellt werden, wenn die Anzahl der Trainingsbilder nur geringfügig erhöht wird.

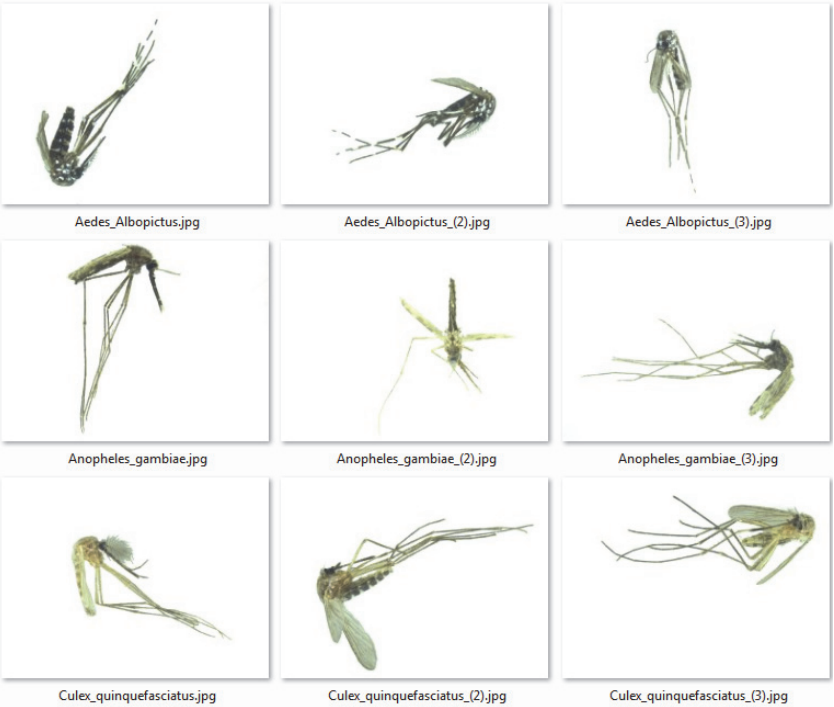


Abbildung 17.3: Trainingsbilder (von oben): Aedes, Anopheles und Culex.

6 Fazit und Ausblick

In dieser Arbeit wurde die Superpixel-gestützte Klassifikation von Stechmückengattungen mit Hilfe der BoF-Methode untersucht. Es wurde eine mögliche Einbettung der Superpixel in die BoF-Methode vorgestellt. Zur Berechnung des Superpixel-Deskriptors wurden drei unterschiedliche Variationen der Superpixeleigenschaften aufgezeigt. Alle 3 Variationen wurden an einer kleinen Reihe von Stechmückenbildern, bestehend aus den Gattungen Aedes, Anopheles und Culex, miteinander und mit den SIFT-Merkmalen verglichen. Es konnte gezeigt werden, dass die Superpixel in manchen Fällen bessere Erkennung von Stechmücken ermöglichen, als die Verwendung der etablierten SIFT-

Merkmale. Die beste Erkennung lieferte der Superpixeldeskriptor, der aus den RGB-Werten und der Größe des Superpixels bestand.

Der Hintergrund wurde sowohl bei den Trainings- als auch bei den Testbildern manuell entfernt. Dies kommt für eine vollautomatische Erkennung nicht in Frage. In Zukunft soll ein Verfahren implementiert werden, das den Hintergrund ähnlich wie in [16] automatisch vom Objekt trennt.

Die Untersuchungen wurden an einer kleinen Anzahl von Referenzbildern durchgeführt. Um zu bewerten, ob die Superpixel als Merkmale für die BoF-Methode in Frage kommen, wird eine größere Anzahl an Referenzbildern benötigt. Ferner wurden nur 3 Klassen klassifiziert. Es muss evaluiert werden, ob die vorgestellte Methode auch bei einer größeren Anzahl an Klassen hohe Trefferquoten liefert.

Literatur

1. P. Reiter, „Climate Change and Mosquito-Borne Disease“, in *Environmental Health Perspectives 109*, 2001, S. 141–161.
2. D. G. Lowe, „Distinctive image features from scale-invariant keypoints“, *Int. J. Comput. Vision*, Vol. 60, Nr. 2, S. 91–110, Nov. 2004.
3. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua und S. Suesstrunk, „SLIC Superpixels Compared to State-of-the-Art Superpixel Methods“, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, Nr. 11, S. 2274–2282, 2012.
4. C. Cortes und V. Vapnik, „Support-vector networks“, *Mach. Learn.*, Vol. 20, Nr. 3, S. 273–297, Sep. 1995.
5. Z. Wu, Z. Hu und Q. Fan, „Superpixel-based unsupervised change detection using multi-dimensional change vector analysis and svm-based classification“, *AnnalsPRS*, Vol. I-7, Nr. 2012, S. 257–262, 2012.
6. B. Fulkerson, A. Vedaldi und S. Soatto, „Class segmentation and object localization with superpixel neighborhoods“, in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2009.
7. A. Lucchi, K. Smith, R. Achanta, V. Lepetit und P. Fua, „A fully automated approach to segmentation of irregularly shaped cellular structures in em images.“ in *MICCAI (2)*, Ser. Lecture Notes in Computer Science, Vol. 6362. Springer, 2010, S. 463–471.

8. J. Cheng, L. Jiang, Y. Xu, F. Yin, D. W. K. Wong, N. M. Tan, D. Tao, C. Y. Cheng, T. Aung und T. Y. Wong, „Superpixel classification based optic disc and optic cup segmentation for glaucoma screening." *IEEE Trans. Med. Imaging*, Vol. 32, Nr. 6, S. 1019–1032, 2013.
9. N. Larios, H. Deng, W. Zhang, M. Sarpola, J. Yuen, R. Paasch, A. Moldenke, D. Lytle, S. Correa, E. Mortensen, L. Shapiro und T. Dietterich, „Automated Insect Identification through Concatenated Histograms of Local Appearance Features", in *IEEE Workshop on Applications of Computer Vision*, 2007, S. 26–26.
10. G. Martinez-Munoz, N. Larios, E. Mortensen, W. Zhang, A. Yamamuro, R. Paasch, N. Payet, D. Lytle, L. Shapiro, S. Todorovic, A. Moldenke und T. Dietterich, „Dictionary-free categorization of very similar objects via stacked evidence trees", in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, S. 549–556.
11. N. Larios, B. Soran, L. Shapiro, G. Martinez-Munoz, J. Lin und T. Dietterich, „Haar Random Forest Features and SVM Spatial Matching Kernel for Stonefly Species Identification", in *20th International Conference on Pattern Recognition (ICPR)*, 2010, S. 2624–2627.
12. Y. Gao, H. Song, X. Tian und Y. Chen, „Identification Algorithm of Winged Insects Based on Hybrid Moment Invariants", in *The 1st International Conference on Bioinformatics and Biomedical Engineering*, 2007, S. 531–534.
13. H. Yang, W. Liu, K. Xing, J. Qiao, X. Wang, L. Gao und Z. Shen, „Research on Insect Identification Based on Pattern Recognition Technology", in *Sixth International Conference on Natural Computation (ICNC)*, Vol. 2, 2010, S. 545–548.
14. G. Csurka, C. Bray, C. Dance und L. Fan, „Visual categorization with bags of keypoints", *Workshop on Statistical Learning in Computer Vision, ECCV*, S. 1–22, 2004.
15. S. Lloyd, „Least squares quantization in pcm", *IEEE Transactions on Information Theory*, Vol. 28, Nr. 2, S. 129–137, Mar. 1982.
16. Z. Ren, S. Gao, L.-T. Chia und I. Tsang, „Region-based saliency detection and its application in object recognition", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. PP, Nr. 99, S. 1–1, 2013.

Zweistufige Anwendung der Saliency-Methodik zur Stechmückendetektion

Jonas Jäger, Viviane Wolff und Klaus Fricke-Neuderth

Hochschule Fulda, Fachbereich Elektrotechnik und Informationstechnik,
Marquardstr. 35, D-36039 Fulda

Zusammenfassung Für bildhafte Detektionsverfahren ist es i. d. R. erforderlich, das zu detektierende Objekt vom Bildhintergrund zu separieren. Im vorliegenden Artikel stellen diese Objekte Stechmücken dar und zur Objektseparierung wird ein modifiziertes *Saliency*-Verfahren entwickelt. Hierzu wird ein bestehender *Saliency*-Algorithmus SF [1] angepasst und mit Hilfe eines zweistufigen Prinzips beschleunigt. Der Grundgedanke des Verfahrens ist es, zunächst eine schnelle Schätzung der Hintergrundpixel durchzuführen, um auf dieser Basis in der zweiten Stufe die exakte Trennung von Objekt und Hintergrund zu erreichen. Die Evaluation erfolgt anhand des umfangreichen und in der Literatur häufig verwendeten *MSRA-1000*-Datensatzes. Die Bewertung des entwickelten Algorithmus ergibt eine Minderung der Ausführungszeit um $29\,\%$ bei gleicher Genauigkeit gegenüber dem Originalalgorithmus SF.

1 Einleitung

Das menschliche Auge kann visuelle Informationen mit hoher Geschwindigkeit verarbeiten. Hierdurch ist der Mensch befähigt, auffällige Objekte einer Szene sicher und schnell zu erkennen. Als auffälliges Objekt wird ein Objekt in einem Bild verstanden, welches vom menschlichen Betrachter beim ersten Anblick als solches erkannt wird. Das Identifizieren von diesen Objekten ist in vielen Anwendungen im Bereich der *Computer-Vision* erforderlich oder wünschenswert. Hierbei sei zum Beispiel an die Bildsegmentierung, die Objektbeschreibung mit Deskriptoren und die Klassifikation von Objekten gedacht.

In der vorliegenden Arbeit werden Bilder unterschiedlicher Stechmückenarten betrachtet. Das Ziel ist es, die korrekte Zuordnung der Bildpixel zu einem Objekt oder dem Hintergrund herzustellen, um darüber auffällige Objekte in den Bildern zu identifizieren (*Saliency-Detection*). Deshalb findet ein zweistufiges Prinzip Verwendung, wobei auf eine Verringerung der Ausführungszeit eines bestehenden *Saliency*-Algorithmus abgezielt wird. Der Grundgedanke des Prinzips besteht darin, in der ersten Stufe eine schnelle Schätzung der Hintergrundpixel durchzuführen. Mit diesem Ergebnis soll in einer zweiten Stufe eine exakte Trennung von Objekt und Hintergrund erreicht werden.

In dieser Untersuchung sollen die Ergebnisse des Forschungsprojekts *Mückenscanner* der Hochschule Fulda weiterverfolgt werden. Hier sollte ein kameragestütztes eingebettetes System entwickelt werden, welches Stechmückenarten anhand von Bilddaten klassifiziert. Um eine solche Klassifikation durchführen zu können, werden Merkmale der Stechmücke extrahiert. Hierzu muss bekannt sein, welche Pixel des Bildes einer Stechmücke zugeordnet sind. Bisher wurde dieser Schritt manuell durchgeführt, indem alle störenden Elemente und der Bildhintergrund mit einem Bildverarbeitungsprogramm entfernt wurden. In dieser Arbeit soll ein Verfahren vorgestellt werden, welches die der Mücke zugeordneten Pixel erkennt, um eine Klassifikation vollkommen automatisch durchführen zu können.

2 Stand der Forschung

In die Detektion von auffälligen Objekten (*Saliency-Detection*) anhand von Bilddaten fanden bereits einige Forschungsarbeiten statt. Eine der ersten Arbeiten auf diesem Gebiet stammt von Itti et al. [2]. Hier wird ein biologisch inspiriertes Modell vorgeschlagen, welches Aufmersamkeits-Karten (*Saliency-Maps*) anhand von Farb-, Helligkeits- und Richtungsmerkmalen mit Hilfe eines *Center-Surround*-Operators über verschieden skalierte Bildkopien mit anschließender Normalisierung berechnet. Es folgten weitere Untersuchungen [3–6], die das *Center-Surround*-Konzept aufgreifen. Dieses Konzept basiert auf der Erkenntnis, dass visuelle Neuronen des Menschen die größte Empfindlichkeit in einer kleinen Region (*Center*) des Blickfeldes aufweisen. Stimuli, die sich weiter vom Zentrum entfernt in einer konzentrischen Um-

gebung (*Surround*) befinden, hemmen hingegen diese neuronale Reaktion.

In neueren Studien ist ein Trend zur Abstraktion als Vorverarbeitungsschritt zu erkennen. Während dieses Schrittes wird das Bild in für die Wahrnehmung homogene Elemente aufgeteilt, was zu einer Unterdrückung von unwesentlichen Details führt. Hierdurch sollen die Objektgrenzen besser erfasst werden. Besonders häufig werden für eine solche Segmentierung *Superpixel*-Verfahren eingesetzt. So wird die Eigenschaft ausgenutzt, dass Pixel, die einem *Superpixel* zugeordnet sind, eine große Ähnlichkeit aufweisen [7] und in der Regel zum selben Objekt gehören.

Wei et al. [8] empfehlen, das Hauptaugenmerk bei der *Saliency*-Bewertung auf den Hintergrund und nicht, wie sonst üblich, auf das hervorstechende Objekt zu legen. Die Methode der Autoren beruht auf einer Abstraktion des Bildes durch Segmentierung. Zur Erstellung der *Saliency-Map* wird wie folgt vorgegangen: Zunächst wird ein 400×400-Pixel großes Bild entweder in 10×10-Pixel Segmente unterteilt (rasterbasiert) oder mit Hilfe des in [9] vorgestellten Superpixelalgorithmus segmentiert (*superpixel*-basiert). Aus diesen Segmenten lässt sich anschließend ein ungerichteter Graph aufbauen. Der *Saliency*-Wert für jedes Segment wird in einem weiteren Schritt aufgrund des kürzesten Pfades zum *Background-Knoten* berechnet. Der *Background-Knoten* ist als ein *virtueller* Knoten des Graphen definiert, der mit allen Segmenten verbunden ist, die den Bildrand berühren und dem Hintergrund zugeordnet sind. Das Gewicht einer Kante zwischen zwei Knoten wird aufgrund des Farbunterschiedes bemessen. Eine Evaluation ließ sich auf Basis der *MSRA-1000* und *Berkeley-300* Datasets durchführen. Die Ausführungszeit des rasterbasierten Algorithmus ist mit durchschnittlich 2 ms (Testhardware: Intel 2.33 GHz CPU, 4 GB Ram) sehr kurz.

Eine weitere *superpixel*-basierte Methode stellen Perazzi et al. [1] vor. Der erste Schritt ihres Verfahrens ist die Abstraktion des Bildes mit Hilfe des SLIC-Algorithmus [7], der Pixel mit ähnlichen Eigenschaften in einer lokalen Umgebung zu Clustern zusammenfasst. Darauf folgend wird der *Einzigartigkeitswert* eines Elements (Clusters) errechnet. Dieser Wert steht mit der Annahme in Verbindung, dass Regionen, die sich in gewissen Aspekten von dem Großteil der anderen Regionen unterscheiden, die Aufmerksamkeit eines menschlichen Betrachters auf sich ziehen. Als weiteres Maß für die *Saliency*-Bewertung wählen die Au-

toren die räumliche Verteilung von Elementen (*Superpixeln*). Dies wird durch die Beobachtung gerechtfertigt, dass Farben, die dem Bildhintergrund angehören in der Regel über das ganze Bild verteilt sind. *Saliency*-Objekte hingegen sind meistens räumlich eher kompakt. In einem letzten Schritt werden die beiden erwähnten Werte zu einem einzigen *Saliency*-Wert kombiniert und pro *Superpixel* festgelegt. Der *Saliency*-Wert des *Superpixel* wird abschließend auf Pixelebene umgerechnet, um eine *Saliency*-Karte mit der Auflösung des Originalbildes zu erhalten.

Wang et al. [10] verwenden in ihrer Arbeit einen zweistufigen Ansatz zur *Saliency*-Detektion. Die erste Stufe beinhaltet eine schnelle, aber grobe *Saliency*-Bewertung. In der zweiten Stufe werden nur noch Teile des Gesamtbildes genauer untersucht. Diese Vorgehensweise soll den menschlichen Wahrnehmungsapparat nachahmen. Psychologische Untersuchungen geben Hinweise darauf, dass auch die Wahrnehmung des Menschen Bilder in einem solchen zweistufigen Prozess verarbeitet.

3 Methode

Für die Identifizierung von Stechmücken werden Bilder mit mittlerer bis hoher Auflösung ($\geq 800 \times 600$ Pixel) verwendet. Ein großer Teil der Bildpixel gehört dabei dem Hintergrund an. Diese Tatsache soll ausgenutzt werden, um die Objekterkennung zu beschleunigen. Deshalb wird zunächst eine grobe *Saliency-Map* mit Hilfe des schnellen *Geodesic-Saliency*-Algorithmus [8] erzeugt. Da eine schnelle Objekterkennung immer auf Kosten der Genauigkeit stattfindet, wird in einer zweiten Stufe eine Schärfung der Objektumrisse mit dem *superpixel*-basierten *SF*-Verfahren [1] durchgeführt. Eine Verkürzung der Ausführungszeit des *SF*-Verfahrens erreicht man durch das Auslassen der bereits als Hintergrund bekannten Pixel während der Berechnung.

3.1 Vorverarbeitung (GS_{GD})

Das GS_{GD}-Verfahren von Wei et al. [8] basiert auf der Grundidee, den Hintergrund in einem Bild zu finden. Auf diese Weise sollen Hintergrund und Objekte mit Hilfe von *Saliency*-Karten getrennt werden. Hierzu werden zwei Annahmen über den Bildhintergrund in natürlichen Szenen getroffen: Zum einen der Gesichtspunkt, dass Be-

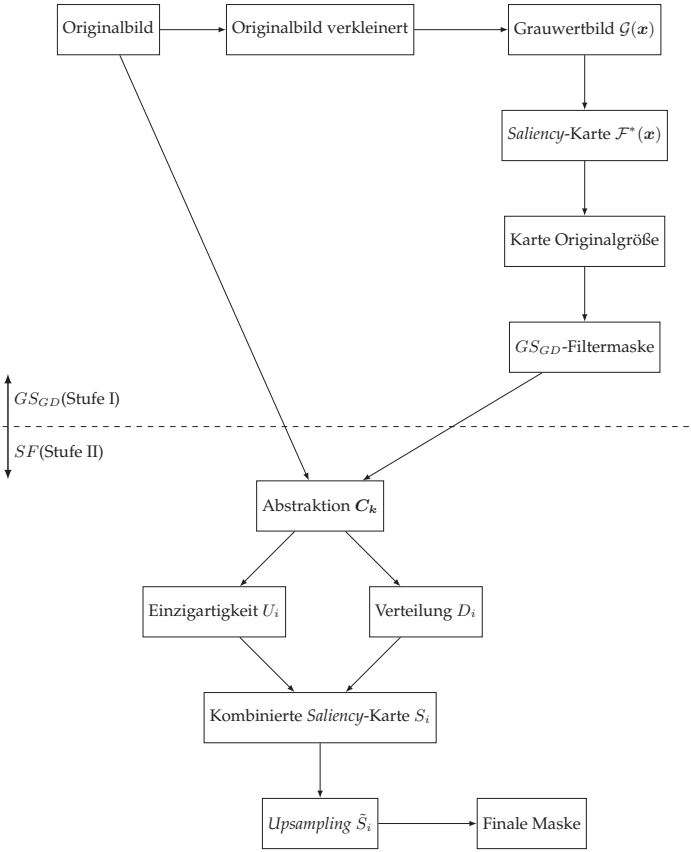


Abbildung 18.1: Darstellung der Stufen und Etappen des TSS-Algorithmus.

reiche, die den Bildrand berühren, wahrscheinlich zum Hintergrund gehören. Dies folgt nach Wei aus einer Regel der Fotografie, die besagt, dass Objekte, die den Fotografen interessieren, nicht den Bildrand schneiden sollten. Zum anderen wird von der Hypothese ausgegangen, dass Hintergrundregionen gewöhnlich groß und homogen sind. Hieraus folgt: Bildelemente, die dem Hintergrund angehören, sollten sich leicht verbinden lassen.

Für die Berechnung von *Saliency*-Karten aufgrund der zwei zuvor erwähnten Annahmen adaptieren Wei et al. die *Geodesic*-Distanz-Transformation von Toivanen [11]. Bei dieser Transformation wird für jedes Pixel der kürzeste Pfad zum nächsten Hintergrund-Pixel berechnet. Somit sind Pixel, die dem Hintergrund angehören, durch einen kleinen Distanzwert gekennzeichnet und solche, die einem *Saliency*-Objekt zugeordnet sind, durch einen großen.

Small-weight-accumulation Bei der einfachen Anwendung der *Geodesic*-Distanz-Transformation besteht das Problem, dass sehr kleine Distanzwerte zwischen zwei Pixeln in der *Saliency*-Karte aufsummiert werden. Hieraus resultiert ein schlechteres Ergebnis. Sehr kleine Distanzwerte zwischen Pixeln in einer bestimmten Region weisen darauf hin, dass alle Pixel der Region einander ähnlich sind und vermutlich zur selben Einheit gehören. Diese Überlegung veranlasst Wei dazu, alle Distanzwerte unterhalb eines bestimmten Schwellwerts auf null zu setzen.

Boundary-edge-weight Ein weiteres Problem stellt der Initialisierungsschritt dar, bei dem festgelegt werden muss welche Pixel, die den Bildrand berühren, dem Hintergrund zugeordnet sind. Es kann nicht davon ausgegangen werden, dass alle Pixel des Bildrandes dem Hintergrund angehören. Um diesem Problem zu begegnen, werden alle Bildrand-Pixel als eindimensionales „Bild“ aufgefasst. Aus diesem „Bild“ wird anschließend mit Hilfe einer einfachen Kontrastmessung [12] eine *Saliency*-Karte erstellt. So kann jedem Bildrand-Pixel ein Startwert zugewiesen werden, der angibt, mit welcher Wahrscheinlichkeit dieser Pixel dem Hintergrund angehört.

GS_{GD}-Filtermaske Den letzten Schritt der Vorverarbeitung stellt die Ermittlung einer Filtermaske dar. Mit Hilfe dieser Maske werden die dem Hintergrund zugehörigen Bereiche markiert. Diese Bereiche muss man bei der anschließenden Verfeinerung (siehe Abschnitt 3.2) nicht mehr beachten, wodurch sich Ausführungszeit einsparen lässt.

3.2 Verfeinerung (SF)

Abstraktion Zunächst sehen Perazzi et al. [1] einen Abstraktionsschritt vor, bei dem das Bild in Basiselemente zerlegt wird – die wesentlichen Merkmale des Bildes bleiben erhalten. Hierfür wird der *Superpixel*-Algorithmus SLIC [7] verwendet. Eine solche Segmentierung ist sinnvoll, um unwesentliche Details für eine spätere Verarbeitung zu entfernen. Außerdem wird hierdurch die Anzahl der zu durchlaufenden Elemente und somit auch die Ausführungszeit der folgenden Schritte verringert. Alle Pixel eines Basiselements besitzen ähnliche Eigenschaften (in diesem Fall die Farbwerte) und würden auch vom menschlichen Betrachter als homogen wahrgenommen. Auffallende Eigenschaften, wie die Konturen des Bildes, bleiben durch eine Anpassung der Basiselemente erhalten. Größe und Form aller Elemente sind ähnlich.

Perazzi et al. bemerken, dass ihr Abstraktionsschritt ca. 40% der Gesamtzeit des SF-Verfahrens in Anspruch nimmt. An dieser Stelle findet die Kombination von GS_{GD} und SF-Algorithmus statt. Hierzu werden alle Pixel, die bereits als Hintergrund-Pixel bekannt sind, bei der Berechnung der Superpixel übersprungen.

Superpixel-Einzigartigkeit Nach Fertigstellung der Abstraktion, wird in einem nächsten Schritt jedem Basiselement ein Einzigartigkeitswert zugewiesen. Die Berechnung der Einzigartigkeitswerte basiert auf der Annahme, dass *Saliency*-Objekte einen deutlichen Unterschied zu Ihrer Umgebung darstellen. Elemente, die einen großen Kontrast zu ihrer Umgebung haben, sollten also einen hohen *Saliency*-Wert besitzen. Hierzu wird die Unterschiedlichkeit eines Elements zu allen anderen Basiselementen, welche das Bild repräsentieren, ermittelt.

Superpixel-Verteilung Die Verteilungswerte verwendet man als zweiten Indikator für die Auffälligkeit eines Objektes. Für gewöhnlich sind Objekte im Vordergrund eher kompakt, wobei der Hintergrund zumeist über das gesamte Bild verteilt ist. Diese Tatsache wird ausgenutzt, um eine Verteilungsmessung zu definieren. Hierbei ordnet man Elementen, die sich nur in einer bestimmten Bildregion befinden, einen großen *Saliency*-Wert zu. Elemente, die hingegen über das gesamte Bild verteilt sind, wird ein geringer Wert zugewiesen.

***Saliency*-Zuweisung** In diesem Schritt erzeugt man die finale *Saliency*-Karte. Hier lassen sich zunächst die beiden bereits erwähnten Kontrastmessungen zu einer einzigen Karte kombinieren. Da dies eine Karte auf Element-Ebene ist, wird sie an dieser Stelle unter Verwendung eines *Upsampling*-Verfahrens [13] auf Pixel-Ebene umgerechnet.

Filtermaske Unter Verwendung der finalen *Saliency*-Karte, die jedem Pixel einen *Saliency*-Wert zuordnet, erzeugt man nun eine Filtermaske. Hierzu werden alle Pixel, die nicht einem *Saliency*-Objekt angehören, mit Hilfe eines adaptiven Schwellwertes ausgeblendet beziehungsweise auf einen festen Wert gesetzt. Daraus resultiert eine binäre Maske, mit der es möglich ist, das auffällige Objekt aus dem Originalbild auszuschneiden.

4 Resultate

4.1 Ausführungsgeschwindigkeit

Die Entwicklung des vorgestellten zweistufigen *Saliency*-Verfahrens TSS hatte das Ziel, einen deutlichen Geschwindigkeitsgewinn ge-

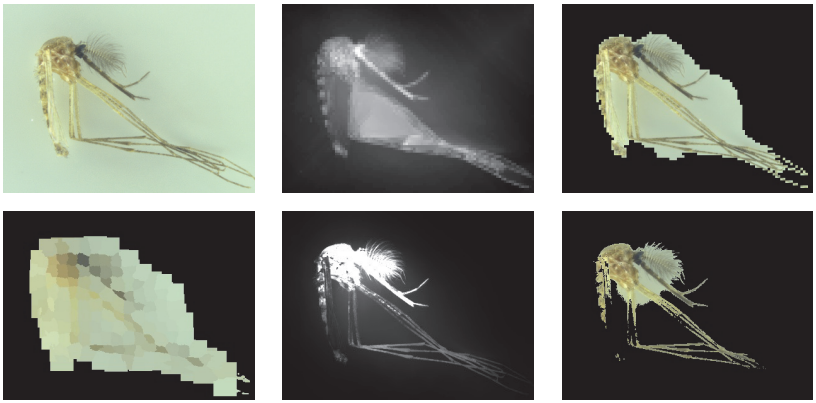


Abbildung 18.2: Ergebnisbilder des vorgestellten TSS-Verfahrens. Von oben links nach unten rechts: (a) Originalbild, (b) GS_{GD}-*Saliency-Map*, (c) GS_{GD}-Filtermaske, (d) Abstraktion, (e) TSS-*Saliency-Map*, (f) Finale Maske.

genüber dem SF-Verfahren zu erreichen. Als Überprüfung wurden die Ausführungszeiten der beiden Algorithmen sowie von GS_{GD} gemessen. Für Testbilder kam die in der einschlägigen Literatur häufig verwendete Untermenge des *MRSA*-Datensatzes zum Einsatz. Diese Untermenge (*MSRA-1000*) umfasst 1000 Bilder des *MSRA*-Datensatzes, zu denen Achanta et al. [5] die zugehörigen *Ground-Truth*-Masken zur Verfügung stellen. Um aussagekräftige Ergebnisse zu erhalten, wurde die durchschnittliche Ausführungszeit, gemittelt über alle 1000 Bilder, für jeden Algorithmus berechnet. Die Tests wurden auf einem Intel Xeon W3690 Prozessor mit $3,46$GHz durchgeführt. Als Arbeitsspeicher standen 12GB RAM zur Verfügung sowie eine Microsoft Windows 7-Umgebung, in der in C++ programmiert und mit Hilfe von Microsoft Visual Studio 2013 im Release-Modus kompiliert wurde.

Die Testergebnisse aus Tabelle 18.1 zeigen, dass die Kombination von GS_{GD} und SF zu TSS eine Verkürzung der Ausführungszeit um 29% zur Folge hat. Die Tests ergeben weiterhin, dass der Vorverarbeitungsschritt des TSS-Verfahrens mit Hilfe des GS_{GD}-Verfahrens durchschnittlich nur 1.92 Millisekunden von insgesamt 115.3 Millisekunden in Anspruch nimmt. Der Vorverarbeitungsschritt benötigt also nur 1.7% der gesamten Ausführungszeit von TSS. Dies legt die Vermutung nahe, dass eine zweistufige Herangehensweise unter Verwendung des GS_{GD}-Algorithmus auch andere *Saliency*-Verfahren beschleunigen könnte.

GS_{GD} [8]	SF [1]	TSS
1.92	162.30	115.30

Tabelle 18.1: Testergebnisse: Durchschnittliche Ausführungszeit in Millisekunden.

4.2 Genauigkeit

Die Tests zur Genauigkeit wurden wieder mit Hilfe des *MSRA-1000* Datensatzes durchgeführt. Abbildung 18.3 zeigt die *Precision-Recall*-Graphen des GS_{GD}-, SF- und TSS-Verfahrens. Es ist deutlich zu erkennen, dass die Genauigkeit von GS_{GD} weitaus geringer ist als die der anderen beiden Algorithmen. Die Graphen von SF und TSS sind fast deckungsgleich. Sie besitzen also die gleiche Genauigkeit nach der *Precision-Recall*-Metrik.

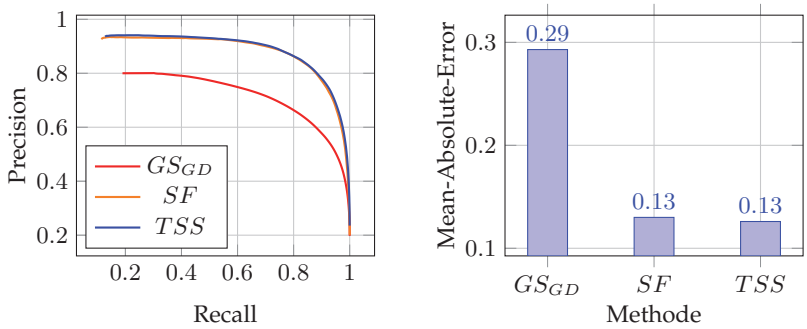


Abbildung 18.3: *Precision-Recall*-Graph und *Mean-Absolute-Error* der drei Verfahren GS_{GD} [8], SF [1] und TSS. Die Ergebnisse wurden über 1000 Bilder des *MSRA*-Datensatzes gemittelt.

In Abbildung 18.3 ist der durchschnittliche *Mean-Absolute-Error* der drei Verfahren zu sehen. Die Werte bestätigen die Beobachtungen, welche schon anhand der *Precision-Recall*-Graphen gemacht werden konnten (s. o.).

5 Fazit

Ausgehend von der Motivation, Stechmücken in hochauflösenden Fotoaufnahmen von ihrem Bildhintergrund zu trennen, um einen Beitrag zu einer vollautomatischen Stechmückenerkennung zu leisten, sollte ein geeigneter *Saliency*-Algorithmus entwickelt werden.

Der Grundgedanke des vorgestellten TSS-Verfahrens bezog sich zunächst darauf, eine schnelle Schätzung der Hintergrundpixel durchzuführen und auf dieser Grundlage in der zweiten Stufe eine exakte Trennung von Objekt und Hintergrund zu erreichen. Auf diese Weise sollte der genauere und zeitintensivere Algorithmus der zweiten Stufe beschleunigt werden, da so die schon bekannten Hintergrundpixel nicht mehr in die Berechnungen einbezogen werden müssen.

Während der anfänglichen Literaturrecherche fiel der GS_{GD}-Algorithmus von Wei et al. [8] als besonders schnelles *Saliency*-Verfahren mit einer sehr kurzen Ausführungszeit auf. Dieser Algorithmus war damit

prädestiniert für die erste Stufe des Verfahrens. Für die zweite Verarbeitungsstufe wurde der SF-Algorithmus von Perazzi et al. [1] gewählt. Gemessen an den *Precision-Recall*-Graphen und dem *Mean-Absolute-Error* schnitt dieser im Hinblick auf die Genauigkeit und im Vergleich zu anderen aktuellen Methoden sehr gut ab. Verglichen mit anderen *genauen* Verfahren konnte er auch hinsichtlich der Ausführungszeit überzeugen.

Ziel des Entwurfs und der Implementierung des zweistufigen Verfahrens war es, den SF-Algorithmus bei gleichbleibender Genauigkeit deutlich zu beschleunigen. Wie die Evaluation anhand des *MSRA*-Datensatzes und an Bildern von Stechmücken zeigt, konnte diese Zielsetzung erreicht werden. Das kombinierte Verfahren dieser Arbeit TSS bearbeitet Bilder bei gleicher Genauigkeit um 29% schneller als der SF-Algorithmus. Die Vorverarbeitungsstufe des TSS-Verfahrens benötigt hierbei nur 1.7% der Gesamtausführungszeit. Es ist anzunehmen, dass auch andere *Saliency*-Verfahren unter Verwendung des hier vorgestellten zweistufigen Prinzips merklich beschleunigt werden können. Der GS_{GD}-Algorithmus stellt hierbei eine hervorragende Wahl für die erste Verarbeitungsstufe dar. Für den Grad der Beschleunigung ist jedoch das Verhältnis von Pixeln des *Saliency*-Objektes zu den Pixeln des Hintergrundes entscheidend. Je größer dieses Verhältnis im betrachteten Bild ist, umso geringer ist die zu erwartende Beschleunigung.

Literatur

1. A. Hornung, Y. Pritch, P. Krahenbuhl und F. Perazzi, „Saliency filters: Contrast based filtering for salient region detection“, *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 0, S. 733–740, 2012.
2. L. Itti, C. Koch und E. Niebur, „A model of saliency-based visual attention for rapid scene analysis“, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 20, Nr. 11, S. 1254–1259, Nov. 1998. [Online]. Available: http://dx.doi.org/10.1109/34.730558
3. S. Frintrop, M. Klodt und E. Rome, „A real-time visual attention system using integral images“, in *In Proc. of the 5th International Conference on Computer Vision Systems (ICVS)*, 2007.
4. D. Gao, V. Mahadevan und N. Vasconcelos, „The discriminant center-surround hypothesis for bottom-up saliency.“ in *NIPS*, J. C. Platt,

D. Koller, Y. Singer und S. T. Roweis, Hrsg. Curran Associates, Inc., 2007. [Online]. Available: http://dblp.uni-trier.de/db/conf/nips/nips2007.html#GaoMV07

5. R. Achanta, S. Hemami, F. Estrada und S. Süsstrunk, „Frequency-tuned Salient Region Detection“, in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, 2009, S. 1597 – 1604, for code and supplementary material, click on the url below. [Online]. Available: http://ivrg.epfl.ch/supplementary_material/RK_CVPR09/index.html
6. D. A. Klein und S. Frintrop, „Center-surround divergence of feature statistics for salient object detection.“ in *ICCV*, D. N. Metaxas, L. Quan, A. Sanfeliu und L. J. V. Gool, Hrsg. IEEE, 2011, S. 2214–2219. [Online]. Available: http://dblp.uni-trier.de/db/conf/iccv/iccv2011.html#KleinF11
7. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua und S. Süsstrunk, „Slic superpixels compared to state-of-the-art superpixel methods“, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, Nr. 11, S. 2274–2282, 2012.
8. Y. Wei, F. Wen, W. Zhu und J. Sun, „Geodesic saliency using background priors“, in *Proceedings of the 12th European Conference on Computer Vision - Volume Part III*, Ser. ECCV'12. Berlin, Heidelberg: Springer-Verlag, 2012, S. 29–42. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-33712-3_3
9. O. Veksler, Y. Boykov und P. Mehrani, „Superpixels and supervoxels in an energy optimization framework“, in *Proceedings of the 11th European Conference on Computer Vision: Part V*, Ser. ECCV'10. Berlin, Heidelberg: Springer-Verlag, 2010, S. 211–224. [Online]. Available: http://dl.acm.org/citation.cfm?id=1888150.1888167
10. Z. Wang und B. Li, „A two-stage approach to saliency detection in images.“ in *ICASSP*. IEEE, 2008, S. 965–968. [Online]. Available: http://dblp.uni-trier.de/db/conf/icassp/icassp2008.html#WangL08
11. P. J. Toivanen, „New geodesic distance transforms for gray-scale images“, *Pattern Recogn. Lett.*, Vol. 17, Nr. 5, S. 437–450, May 1996. [Online]. Available: http://dx.doi.org/10.1016/0167-8655(96)00010-4
12. H. Zhang, W. Wang, G. Su und L. Duan, „A simple and effective saliency detection approach.“ in *ICPR*. IEEE, 2012, S. 186–189. [Online]. Available: http://dblp.uni-trier.de/db/conf/icpr/icpr2012.html#ZhangWSD12
13. J. Dolson, J. Baek, C. Plagemann und S. Thrun, „Upsampling range data in dynamic environments.“ in *CVPR*. IEEE, 2010, S. 1141–1148. [Online]. Available: http://dblp.uni-trier.de/db/conf/cvpr/cvpr2010.html#DolsonBPT10

Industrielle Sortierung von Mineralen anhand von hyperspektralen Fluoreszenzaufnahmen – Potenzialbewertung

Sebastian Bauer und Fernando Puente León

Karlsruher Institut für Technologie, Institut für Industrielle Informationstechnik
Hertzstraße 16, Geb. 06.35, 76187 Karlsruhe

Zusammenfassung Da sich viele Minerale farblich kaum unterscheiden, ist eine Untersuchung alternativer optischer Verfahren notwendig. Aus diesem Grund wurden die Fluoreszenzspektren von ausgewählten Mineralen mit der Methode der hyperspektralen Bildgewinnung aufgezeichnet. Dadurch lassen sich die in Laboruntersuchungen anhand von Punktspektren gezogenen Schlussfolgerungen auf die industrielle Sortierung übertragen, denn für jegliche industrielle Anwendung ist es zentral, dass ein ausreichend hoher Durchsatz erreicht wird. Dies lässt sich nur durch die simultane Bewertung vieler Mineralobjekte realisieren. Es werden erste Klassifikationsergebnisse anhand der aufgezeichneten hyperspektralen Aufnahmen präsentiert.

1 Einleitung

Die Unterscheidung von primären Mineralrohstoffen anhand von optischen Kriterien ist eine anspruchsvolle Aufgabe. Aufgrund des geringen farblichen Unterschieds wird oftmals auch das Nahinfrarotspektrum von ca. $1000\,\mathrm{nm} - 2500\,\mathrm{nm}$ betrachtet. In diesem Wellenlängenbereich haben viele Minerale aufgrund von physikalischen Phänomenen charakteristische spektrale Signaturen [1]. Ebenda wurde außerdem die Eignung von Fluoreszenzspektren von Mineralen zur optischen Unterscheidung untersucht, allerdings wurden die Materialproben vorher geschnitten und poliert sowie lediglich deren Punktspektren betrachtet.

Als Fluoreszenz bezeichnet man den Effekt, dass Stoffe bei optischer Anregung eine Strahlung mit anderer, meist längerer, Wellenlänge (so-

genannte Stokes-Verschiebung) abgeben. Dieser Effekt ist bei einzelnen Stoffen oder Mineralen mehr oder weniger stark ausgeprägt und hängt außerdem von der einfallenden Wellenlänge ab [2].

Bei der *hyperspektralen Bildgewinnung* (Hyperspectral Imaging, HSI), wird pro betrachteter Wellenlänge ein vollständiges Bild der Szene aufgezeichnet [3]. Für jedes einzelne Pixel liegt damit ein vollständiges Spektrum vor. Aus diesem Grund ist die hyperspektrale Bildgebung auch als *bildgebende Spektroskopie* bekannt.

Nach unserem Kenntnisstand werden hyperspektrale Fluoreszenzuntersuchungen bisher lediglich in den Lebenswissenschaften und zur Qualitätsbeurteilung von Lebensmitteln [4, 5], nicht aber für die Mineralsortierung durchgeführt. Dies könnte in der niedrigen Emissionsintensität einiger industriell relevanter Minerale begründet sein. Aufgrund der apparativen Ausstattung ist das Institut für Industrielle Informationstechnik in der Lage, hyperspektrale Fluoreszenzbilder von vielen Mineralen zu generieren. Mit einer durchstimmbaren, monochromatischen Lichtquelle werden Minerale mit verschiedenen Wellenlängen angeregt und das Fluoreszenzleuchten mit einem Acousto-Optical Tunable Filter (AOTF) und einer äußerst rauscharmen EMCCD-Kamera aufgezeichnet. Dieser Beitrag teilt sich in folgende Abschnitte auf: Nach der Beschreibung der zugrundeliegenden physikalischen Phänomene in Kapitel 2 wird in Kapitel 3 der Versuchsaufbau erklärt. Kapitel 4 beleuchtet die Eigenschaften der aufgezeichneten Hyperspektralbilder, während Kapitel 5 auf die verwendeten Klassifikatoren sowie deren Training eingeht. Die Zusammenfassung in Kapitel 6 bildet den Abschluss.

2 Physikalische Grundlagen der Fluoreszenz

Fluoreszenz ist ein Phänomen aus der Gruppe der Lumineszenzeffekte. Lumineszenz entsteht, wenn in einem Körper Elektronen aus angeregten Zuständen zurück in den Grundzustand fallen und dabei Photonen ausgesandt werden. Wird die Anregung durch Photonen hervorgerufen, so nennt man die dabei entstehende Strahlung Photolumineszenz. Innerhalb der Photolumineszenz unterscheidet man zwischen der kurzlebigen Fluoreszenz (nach ca. $10^{-7}\,\mathrm{s}$ abgeklungen) und der langlebigen Phosphoreszenz, die auch nach Stunden oder Tagen noch vor-

handen sein kann. Die unterschiedliche Dauer des Abklingens liegt in dem jeweiligen Zustandsübergang begründet [6]. Da aufgrund von internen, strahlungslosen Verlustprozessen der strahlende Übergang bei der Fluoreszenz unabhängig von der Anregungsenergie in den meisten Fällen vom niedrigsten angeregten Zustand ausgeht [6], besitzt das emittierte Licht eine größere Wellenlänge als das Anregungslicht. Durch diesen Wellenlängenunterschied ist es möglich, simultan zu beleuchten und die Fluoreszenzantwort aufzuzeichnen, ohne das Anregungslicht mittels schmalbandiger Filter vor dem Aufnahmeinstrument unterdrücken zu müssen. Für die industrielle Sortierung werden Fluoreszenzeffekte bislang kaum eingesetzt [1]. Außerdem existieren vereinzelte Ansätze [2], die nicht die spektralen Fluoreszenzeigenschaften von Mineralen, sondern den zeitlichen Abfall der Fluoreszenzintensität im Nanosekundenbereich betrachten.

3 Versuchsbeschreibung

3.1 Anregung und Aufnahme

Abbildung 19.1 zeigt den Versuchsaufbau. Dieser befindet sich in einer abgedunkelten Kammer. Als Lichtquelle dient eine $300\,\mathrm{W}$-Xenon-Kurzbogenlampe, aus deren Spektrum der gewünschte Wellenlängenbereich mittels eines $300\,\mathrm{mm}$-Monochromators ausgeschnitten wird. Im verwendeten Czerny-Turner-Monochromator sind Brechungsgitter enthalten, die das Licht in seine Wellenlängen aufspalten [7]. Mit den Daten aus dem Datenblatt ergibt sich eine Halbwertsbreite des auf die Probe eingestrahlten Wellenlängenbands von $15\,\mathrm{nm}$. Als zentrale Wellenlängen wurden $220\,\mathrm{nm}$ bis $400\,\mathrm{nm}$ in Schritten von $20\,\mathrm{nm}$ verwendet.

Mittels eines Spiegels wird das Licht auf die Probe gelenkt. Das von dieser ausgesandte Fluoreszenzlicht wird mit einem Acousto-Optical Tunable Filter (AOTF) gefiltert (Gooch&Housego HSi-300). Ein AOTF besteht im Wesentlichen aus einem optischen Kristall (hier: Tellur-Dioxid), der mit Schallwellen im Radio-Frequenzbereich angeregt wird. Das auf dem Kristall eintreffende Licht wird gebeugt, wobei der interessierende Anteil in die Kamera geleitet wird. Die weitergeleitete Lichtwellenlänge hängt dabei von der Frequenz der Radiowellen ab und kann somit frei eingestellt werden. Der verwendete AOTF ist ein Flächenfilter, das heißt, es wird eine bestimmte Wellenlänge transmit-

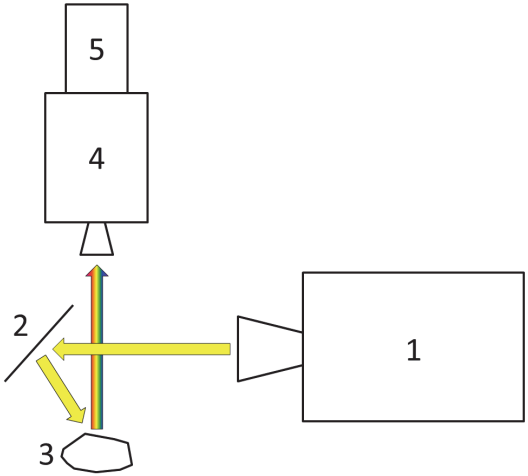


Abbildung 19.1: Versuchsaufbau: Lichtquelle (1), Spiegel (2), Probe (3), AOTF-Spektralfilter (4), EMCCD-Kamera (5).

tiert, ein Bild bei dieser Wellenlänge aufgenommen, die Wellenlänge geändert, das nächste Bild aufgenommen usw. Zur Bildaufnahme wird eine EMCCD-Kamera (Andor iXon$_3$ 897) verwendet. Bei diesem Kameratyp werden die auf dem Chip entstandenen Photoelektronen vor der A/D-Wandlung elektrisch vervielfacht und somit das Ausleserauschen verringert. Der Kamerachip wird auf $-85°\mathrm{C}$ gekühlt, womit der Dunkelstrom vernachlässigbar ist. Besonders bei geringen Lichtintensitäten ist das Gesamtrauschen der EMCCD-Kamera sehr gering. Die Flureszenzantwort der Proben wurde im Wellenlängenbereich von $450\,\mathrm{nm} - 790\,\mathrm{nm}$ aufgezeichnet, wobei der Abstand der Kanalabstand jeweils $17\,\mathrm{nm}$ bei einer Halbwertsbreite von ebenfalls $17\,\mathrm{nm}$ betrug, sodass das gesamte Spektrum erfasst wurde. Die Belichtungszeit betrug pro Wellenlänge $30\,\mathrm{s}$, um genügend Photonen erfassen zu können. Generell wurde weder die Anregung noch die Aufnahme spektral kalibriert. Dadurch lassen sich die Spektren und Ergebnisse nicht auf andere Messaufbauten übertragen, aber da alle in diesem Beitrag beschriebenen Ergebnisse mit dem selben Aufbau gewonnen wurden, sind sie untereinander dennoch vergleichbar.

3.2 Verwendete Minerale

Es wurden hyperspektrale Fluoreszenzbilder von insgesamt 23 oberflächlich gereinigten Mineralen aus vier Sorten (Calcit, Dolomit, Magnesit, Talk) aufgezeichnet. Hierbei stammen die Minerale auch innerhalb der jeweiligen Sorte aus verschiedenen Lagerstätten. Selbst mit bloßem Auge sind zwischen einigen Steinen bereits deutliche Farbunterschiede zu erkennen. Aufgrund der Tatsache, dass sich auch die Fluoreszenzspektren zwischen zwei Mineralen derselben Sorte aus verschiedenen Lagerstätten deutlich unterscheiden, wurden mit Ausnahme eines dunklen Magnesits nur diejenigen Steine aller Klassen berücksichtigt, die weißlich erscheinen, da sich die anderen bereits mit herkömmlichen Farbbildern unterscheiden lassen. Abbildung 19.2 zeigt exemplarisch RGB-Bilder eines Magnesits und eines Dolomits. Diese Bilder verdeutlichen, dass eine einfache Unterscheidung mittels herkömmlicher Farbbilder nicht möglich ist. Die verbliebenen 13 weißlichen Minerale werden in 7 Klassen eingeteilt, um zwischen Mineralen der gleichen Sorte, die aus unterschiedlichen Lagerstätten stammen, unterscheiden zu können. Die Minerale, ihre Klassenzuordnung sowie die in diesem Beitrag verwendete Farbcodierung ist in Tabelle 19.1 angegeben.

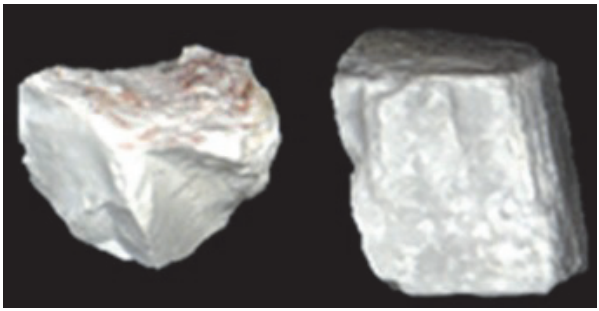


Abbildung 19.2: Weißliche Minerale Magnesit (links) und Dolomit (rechts).

4 Bildeigenschaften

Trotz der langen Belichtungszeit sind nur wenige Photonen erfasst worden. Abbildung 19.3 zeigt im oberen Bildteil das Fluoreszenzbild von

Tabelle 19.1: Verwendete Minerale, ihre Klassenzuordnung und Farbcodierung.

	Klasse	Mineralsorte	Anzahl Steine
1	Magnesit	1	rot
2	Magnesit	3	blau
3	Magnesit	4	grün
4	Magnesit	2	cyan
5	Talk	1	schwarz
6	Dolomit	1	gelb
7	Dolomit	1	magenta

6 Magnesiten bei einer Anregung mit $340\,\mathrm{nm}$ und einer Aufnahme bei $535\,\mathrm{nm}$. Im unteren Bildteil ist eine Aufnahme derselben Szene bei schwachem Tageslicht zu sehen. Wie man sieht, ist der Magnesit rechts unten im Fluoreszenzbild kaum zu sehen. Dies ist der einzelne dunkle Magnesit, der bei der Klassifikation mit berücksichtigt werden soll.

Die Pixelspektren der 7 Klassen sind in Abbildung 19.4 zu sehen. Die Tatsache, dass Spektren des gleichen Steins in y-Richtung um einen bestimmten Faktor gestaucht sind, ist bekannt und liegt in der Tatsache begründet, dass das Fluoreszenzlicht an jedem Punkt diffus strahlt und deswegen die Spektren der Pixel an den Seitenflächen der Steine mit geringerer Intensität registriert werden. Dieser Fakt wird beispielsweise von Kim et al. [4] erwähnt und diskutiert. Die Autoren merken an, dass Richtungseinflüsse Reflektanzbilder eventuell mehr beeinflussen als Fluoreszenzbilder. Diese Richtungsabhängigkeit der aufgezeichneten Spektren hat zur Folge, dass Pixel an den Objekträndern, also schrägen Seitenflächen, oftmals falsch klassifiziert werden. Wie aus Abbildung 19.4 ebenfalls ersichtlich, überlappen die Spektren der verschiedenen Klassen sehr stark. Dieser Effekt unterstreicht die Schwierigkeit der Mineralklassifikation, allerdings berichten Negara et al. [8] im Falle von hyperspektralen Reflektanzbildern von ähnlichen Überlappungen, aber nichtsdestotrotz erfolgreicher Klassifikation. Außerdem lässt sich in Abbildung 19.4 erkennen, dass die Spektren relativ starkem Rauschen unterworfen sind.

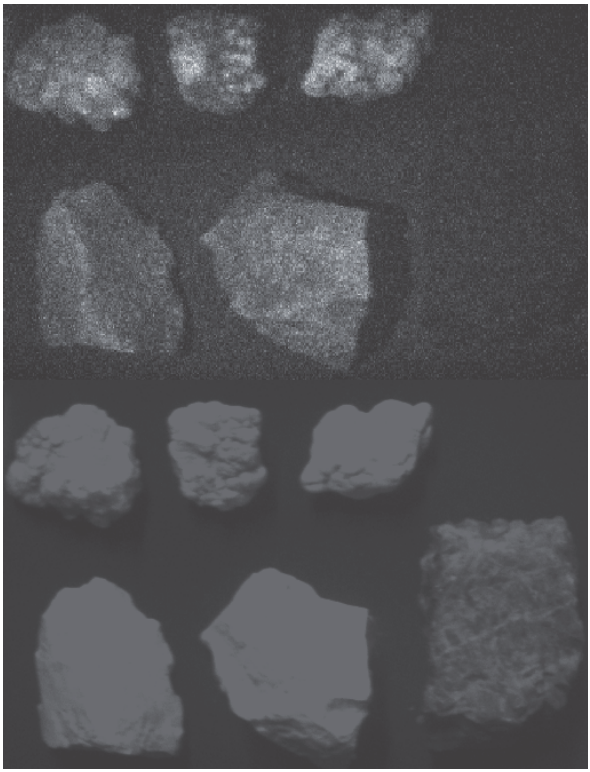


Abbildung 19.3: Oben: Aufnahme von 6 Magnesiten bei 535nm unter Beleuchtung mit 340nm. Unten: Dieselbe Szene bei schwachem Tageslicht.

5 Klassifikation und Ergebnisse

5.1 Klassifikatoren und deren Training

Die Klassifikation der Minerale wurde mit 3 Klassfikatoren durchgeführt: Lineare Diskriminanzanalyse (LDA), Quadratische Diskriminanzanalyse (QDA) und k-nächste-Nachbarn (KNN). Dabei wurde bei LDA und QDA jeweils eine N-fache Kreuzvalidierung durchgeführt. Dafür wurde die Anzahl der verfügbaren Pixel in N Gruppen aufgeteilt und eine Gruppe als Test- und die anderen Gruppen als Trainingsda-

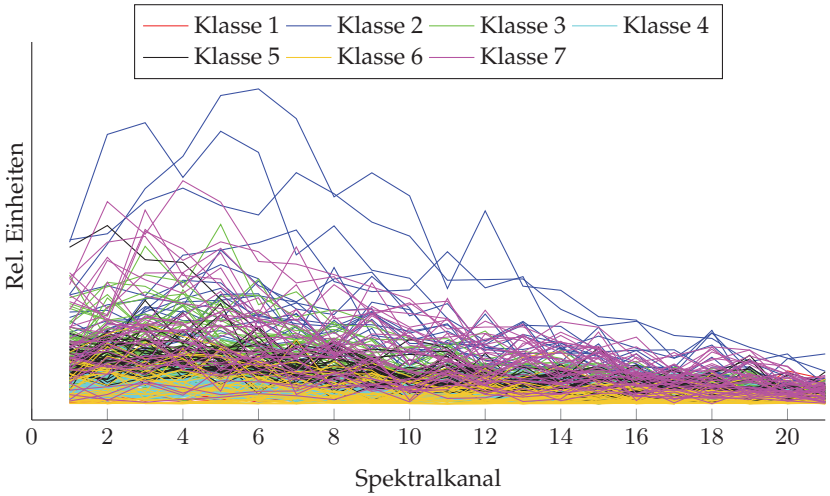


Abbildung 19.4: Spektren der verschiedenen Minerale bei einer Anregung mit Licht der Wellenlänge $340\mathrm{nm}$. Zur besseren Übersichtlichkeit ist nur jedes 1000. Spektrum gezeigt.

ten verwendet. Diese Klassifikation wird N-mal wiederholt, sodass alle Pixel einmal als Testdaten behandelt wurden. Der KNN-Klassfikator wird auf eine andere Art trainiert; da er die Testdaten mit allen vorhandenen Trainingsdaten vergleicht, sollte die Anzahl der aus jeder Klasse als Trainingsdaten verwendeten Pixel gleich sein. Hier sollen aus jeder Klasse Pixelspektren der Anzahl $1/N$ mal der Anzahl der Pixel des kleinsten Steins als Trainingsdaten zur Verfügung stehen. Diese Spektren wurden aus einem Streifen durch die Steinmitte gewonnen, wobei innerhalb des Streifens lediglich jeder N-te Pixel als Trainingsdatum verwendet wird. Jeder Stein trägt je nach Anzahl der in einer Klasse vorhandenen Steine zu der Gesamtzahl der benötigten Trainingspixel bei. Durch dieses Verfahren ist sichergestellt, dass aus jedem Stein Trainingspixel vorhanden sind, dass sowohl Mitten- als auch Randpixel enthalten sind und dass aus jeder Klasse gleich viele Trainingspixel zur Verfügung stehen. Für den KNN wurde k zu 5 gewählt.

Im ersten Schritt werden die Steine mit den unbearbeiteten Pixelspektren klassifiziert. Da die Spektren der Randpixel eines Steins eine gerin-

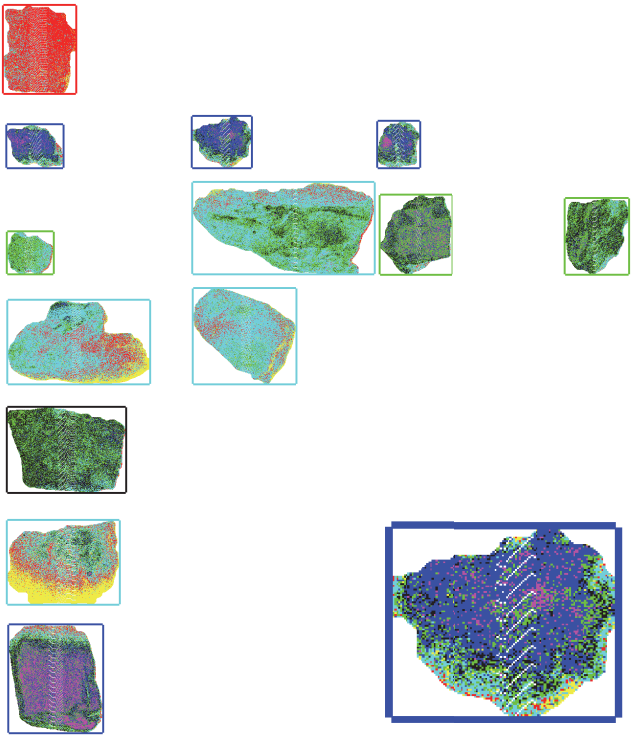


Abbildung 19.5: Pixel- und Objektklassifikationsergebnisse des KNN-Klassifikators bei Anregung mit 340 nm und N=10. Die vergrößerte Darstellung rechts unten veranschaulicht die Wahl der KNN-Trainingspixel.

gere Intensität aufweisen als die Mittenpixel, wird für den zweiten Klassifikationsdurchlauf jedes Pixelspektrum durch seinen Flächeninhalt geteilt und somit normiert. Aufgrund des starken Rauschens werden im dritten Schritt die Pixelspektren vor der Normierung zunächst mit einem Moving-Average-Filter der Länge 3 geglättet.

Da die spektralen Unterschiede zwischen den betrachteten Steinen bei den Anregungswellenlängen 340 nm, 360 nm und 380 nm am größten sind, werden im Folgenden nur die Hyperspektralbilder bei diesen drei Anregungswellenlängen untersucht.

5.2 Klassifikationsergebnisse

Generell wird ein Stein der Klasse zugeordnet, in die die meisten seiner Pixel klassifiziert werden. In Tabelle 19.2 werden die jeweiligen Objektklassfikationsraten vorgestellt. Abbildung 19.5 zeigt exemplarisch die KNN-Klassifikationsergebnisse bei Anregung mit $340\,\mathrm{nm}$ unter Verwendung von $N = 10$. Die Zeilen zeigen die einzelnen Klassen. Die Farbe des Rahmen um jedes Objekt gibt an, in welche Klasse es klassifiziert wurde. Die vergrößerte Abbildung des mittleren Steins von Gruppe 2 in der rechten unteren Bildecke verdeutlicht nochmals die Wahl der Trainingspixel; die weißen Punkte stehen für Pixel, die als Trainingspixel benutzt und deshalb nicht klassifiziert wurden. Es fällt auf, dass, wie bereits beschrieben, vor allem die Randpixel der Steine falsch klassifiziert werden.

5.3 Diskussion

Die Klassifikationsergebnisse hängen stark vom verwendeten Klassifikator ab. Unabhängig von der Vorverabeitung der Spektren liefert KNN fast immer die besten Ergebnisse; bei $360\,\mathrm{nm}$ werden sogar alle Objekte richtig klassifiziert. Es fällt auf, dass die Vorverarbeitung das Klassifikationsergebnis eher verschlechtert als verbessert.

6 Zusammenfassung

Es wurden hyperspektrale Aufnahmen der UV-Fluoreszenz einer kleinen Stichprobe von 23 Mineralen gemacht. Aufgrund der Sensitivität und dem geringen Rauschen der Kamera lassen sich auch niedrigste Fluoreszenzintensitäten bei Mineralen, die mit bloßem Auge nicht als fluoreszierend erkannt werden, registrieren. Es fällt auf, dass die Spektren generell sehr breit sind und sich zwischen den Sorten nur wenig unterscheiden. Diese Eigenschaft erschwert die Wahl geeigneter Wellenlängenbereiche, um beispielsweise mit geschickt gewählten Bandpässen diskriminative Merkmale gewinnen zu können [9].

Die einzelnen Mineralpixel wurden mittels der drei Klassifikatoren LDA, QDA und KNN klassifiziert, wobei KNN die besten Klassifikationsergebnisse liefert. Diese erste Potenzialanalyse zeigt, dass die Klassifikation von Mineralen mittels ihrer Fluoreszenzspektren anspruchs-

Tabelle 19.2: Klassifikationsergebnisse bei Verwendung von 7 Klassen.

	Pixelspektren	**norm. Pixelspektren**	**gef., norm. Pixelspektren**
340nm, N=2			
LDA	0,615	0,385	0,385
QDA	0,615	0,231	0,615
KNN	0,923	0,692	0,692
340nm, N=10			
LDA	0,615	0,385	0,385
QDA	0,615	0,615	0,769
KNN	0,769	0,692	0,692
360nm, N=2			
LDA	0,692	0,538	0,538
QDA	0,692	0,231	0,385
KNN	1	0,846	0,923
360nm, N=10			
LDA	0,692	0,538	0,538
QDA	0,692	0,769	0,462
KNN	1	0,846	0,923
380nm, N=2			
LDA	0,615	0,769	0,769
QDA	0,846	0,846	0,615
KNN	0,923	0,769	0,923
380nm, N=10			
LDA	0,615	0,769	0,769
QDA	0,846	0,462	0,769
KNN	0,923	0,769	0,769

voll, aber prinzipiell möglich ist. Nächste Schritte beinhalten das Untersuchen eines größeren Datensatzes und die Anwendung weiterer Klassifikatoren.

Literatur

1. I. Hofer, R. Huber und K. Weingrill, G. und Gatterer, „Luminescence-and reflection spectroscopy for automatic classification of various minerals“, in *OCM 2013-Optical Characterization of Materials-conference proceedings*. KIT Scientific Publishing, 2013, S. 227.
2. J. Pollmanns, „Identifikation von mineralien und schüttgütern mit hilfe der laserinduzierten fluoreszenz“, *Bergbau*, Vol. 7, S. 322–325, 2008.
3. M. Michelsburg, R. Gruna und F. Vieth, K. und Puente León, „Spektrale bandselektion beim entwurf automatischer sortieranlagen“, in *Forum Bildverarbeitung 2010*. KIT Scientific Publishing, 2010, S. 389–400.
4. M. Kim, Y. Chen, P. Mehl *et al.*, „Hyperspectral reflectance and fluorescence imaging system for food quality and safety“, *Transactions-American Society of Agricultural Engineers*, Vol. 44, Nr. 3, S. 721–730, 2001.
5. H. K. Noh und R. Lu, „Hyperspectral laser-induced fluorescence imaging for assessing apple fruit quality“, *Postharvest Biology and Technology*, Vol. 43, Nr. 2, S. 193–201, 2007.
6. J. Lakowicz, *Principles of Fluorescence Spectroscopy*. Springer, 2007. [Online]. Available: http://books.google.de/books?id=-PSybuLNxcAC
7. http://www.lot-qd.de/files/downloads/lightsources/en/monochromatische-lichtquellen/LQ_Monochromatic_light_sources_ en.pdf.
8. C. Negara, K.-U. Vieth, M. Lafontaine und M. Freund, „Automatic fruit sorting by non-destructive determination of quality parameters using visible-near infrared to improve wine quality: Ii. regression analysis“, *NIR news*, Vol. 25, Nr. 1, S. 4–6, 2014.
9. M. Michelsburg, R. Gruna, K.-U. Vieth und F. Puente León, „Spektrale bandselektion für das filterdesign optischer inspektionssysteme“, *tm-Technisches Messen Plattform für Methoden, Systeme und Anwendungen der Messtechnik*, Vol. 78, Nr. 9, S. 384–390, 2011.

Ellipsometrie an gekrümmten Oberflächen

Christian Negara und Matthias Hartrumpf

Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung
Fraunhoferstraße 1, D-76131 Karlsruhe

Zusammenfassung Die Ellipsometrie ist ein etabliertes und sehr sensitives Messverfahren zur Schichtdickenmessung an ebenen Oberflächen und wird zur stichprobenhaften Qualitätskontrolle vielfach eingesetzt. Durch ein neu entwickeltes Messverfahren können Polarisationsmessungen an gekrümmten Oberflächen mithilfe eines Laserscanners durchgeführt werden, wodurch eine großflächige Oberflächenprüfung im Durchlauf ermöglicht wird. Dieser Artikel behandelt die Gemeinsamkeiten und die Unterschiede zwischen der klassischen Ellipsometrie und der neu entwickelten Retroreflex-Ellipsometrie hinsichtlich der gemessenen physikalischen Größen und der detektierbaren Oberflächen und Schichtdicken.

1 Einleitung

Mit zunehmender Miniaturisierung in mechanischen, optischen und elektrischen Systemen steigt die Notwendigkeit zum Einsatz präziser Messgeräte in der Qualitätssicherung und der Prozesssteuerung bzw. -regelung. Die Vermessung von dünnen Schichten ist bei der Herstellung optischer Elemente, der Herstellung von Verbundstoffen, in der Halbleitertechnik und speziell in der Dünnschichttechnik von elementarem Interesse. Die Ellipsometrie kann bei teilweise reflektierenden und nicht opaken Oberflächen zur Vermessung von Schichtdicken vom Ångström- bis in den Mikrometerbereich eingesetzt werden. Dabei wird ein Lichtstrahl mit definierter Polarisation aus der Sendeeinheit ausgestrahlt und trifft auf die zu prüfende Oberfläche auf. Abhängig von der Anzahl der Schichten, den optischen Materialkonstanten und den Schichtdicken wird der Polarisationszustand des einfallenden Lichts geändert [1]. Durch mehrere optische Elemente wird

der Polarisationszustand des reflektierten Lichts in der Empfangseinheit analysiert (siehe Abb. 20.1).

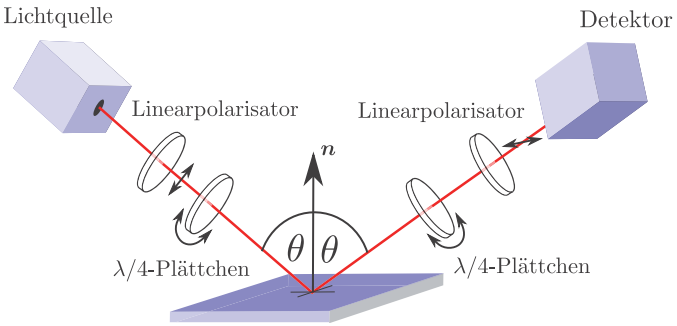
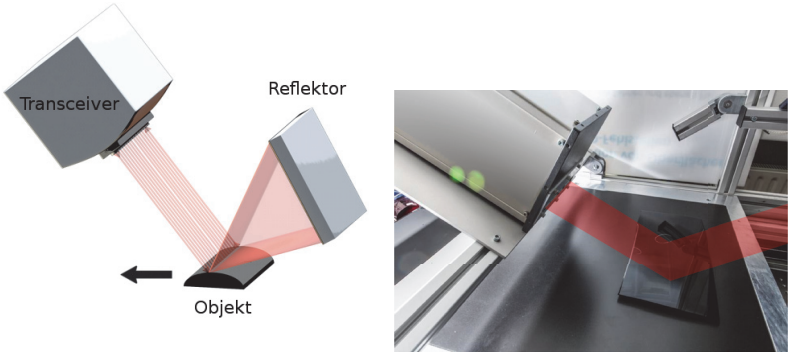


Abbildung 20.1: Schematischer Aufbau eines Ellipsometers mit rotierenden $\lambda/4$-Plättchen und statischen Linearpolarisatoren.

Beim gewöhnlichen ellipsometrischen Aufbau werden die Winkel der Lichtquelle und des Sensors bzgl. der Oberflächennormalen so festgelegt oder eingestellt, dass die Reflexionsbedingung für den Messpunkt oder die Messebene erfüllt wird. Bildgebende ellipsometrische Messgeräte tolerieren keine Neigungsänderung der Oberfläche und stellen hohe Anforderungen an die verwendeten Optiken um parallele Strahlengänge zu gewährleisten [2]. Bei Punktmessgeräten existieren Erweiterungen um kleine Neigungsänderungen von bis zu acht Grad zu tolerieren [3, 4].

2 RRE-Scanner

Ein am Fraunhofer IOSB und von der Firma Opos entwickeltes und patentiertes Messverfahren, die Retroreflex-Ellipsometrie (RRE), erlaubt die großflächige Vermessung von stark gekrümmten Oberflächen mittels Ellipsometrie [5]. Dabei kommt ein Laserscanner mit einer Zeilenfrequenz von 1 kHz zum Einsatz. Ein zirkular polarisierter Laserstrahl tastet die Probenoberfläche linienförmig ab. Das an der Oberfläche der Probe reflektierte Licht wird an einem Retroreflektor auf genau dem gleichen Strahlenweg zurück reflektiert und gelangt so nach nochmaliger Reflexion an der Probenoberfläche in die kombinierte Sende- und



(a) (b)

Abbildung 20.2: Schematische Darstellung des Strahlenverlaufs beim RRE (a) und Bild des Ellipsometriescanners (b).

Empfangseinheit (Abb. 20.2). In der Empfangseinheit wird der Polarisationszustand der reflektierten Strahlung detektiert. Im Gegensatz zum klassischen Ellipsometer ist die tolerierbare Neigungsänderung sehr groß, da der an der Probe reflektierte Strahl lediglich den Reflektor treffen muss.

3 Interaktion von Licht mit Materie

Die folgenden Abschnitte geben einen Überblick über die Beschreibung von Polarisationszuständen von Licht und die Änderung bei der Reflexion an Oberflächen. Es wird ein Zusammenhang zwischen den Messwerten der klassischen Ellipsometrie und denen der Retroreflex-Ellipsometrie hergestellt, so dass bekannte Methoden zur Schichtdickenbestimmung angewendet werden können.

3.1 Polarisation

Licht ist eine transversale elektromagnetische Welle und damit polarisierbar. Polarisiertes Licht ist im allgemeinen Fall elliptisch pola-

risiert. Durch Polarisationsfilter kann es in zueinander orthogonale Zustände zerlegt werden. Eine Möglichkeit ist die Zerlegung in zwei senkrecht zueinander stehenden linear polarisierten Wellen. Die elektrischen Felder der beiden Wellen schwingen in zwei senkrecht zueinander stehenden Schwingungsebenen (hier die x/z bzw. y/z-Ebene), wie in Abb. 20.3 dargestellt. Durch Superposition der beiden elektrischen Feldvektoren erhält man die Schwingung des elektrischen Feldes $\boldsymbol{E}(t)$ der ursprünglichen Welle. Im allgemeinen Fall vollzieht der elektrische Feldvektor $\boldsymbol{E}(t)$ in der Projektion auf die x-y-Ebene eine elliptische Schwingung.

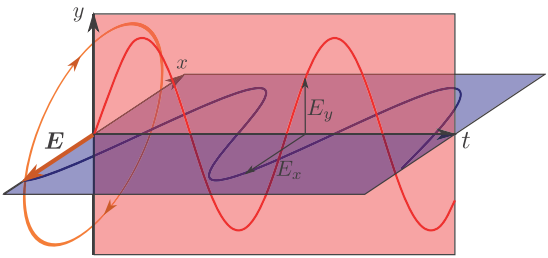


Abbildung 20.3: Darstellung einer elliptisch polarisierten Welle durch Überlagerung von zwei linear polarisierten Wellen.

3.2 Reflexion

Das Reflexionsverhalten von Licht an einer Grenzfläche zwischen zwei Materialien wird durch die Fresnel'schen Formeln beschrieben [6]. Sie setzen Amplitude und Phase der eingestrahlten und reflektierten Strahlung über die komplexen Reflexionskoeffizienten in Beziehung. Die einfallende und die reflektierte Welle werden in zwei linear polarisierte Wellen zerlegt, deren elektrische Feldvektoren E^s, E^p senkrecht und parallel zur Einfallsebene schwingen. E_i^s, E_i^p seien die Amplituden der einfallenden und E_r^s, E_r^p die Amplituden der reflektierten Strahlung. Die Einfallsebene wird durch die Oberflächennormale und die Einstrahlrichtung aufgespannt (Abb. 20.4). Die Reflexionskoeffizienten r_{12}^p, r_{12}^s sind definiert als:

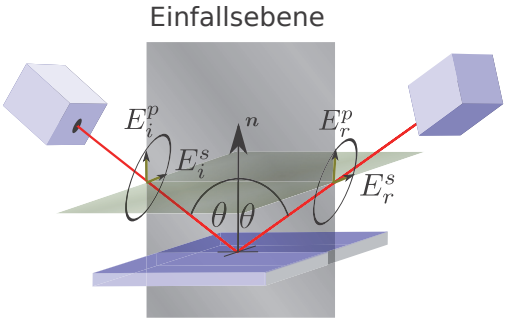


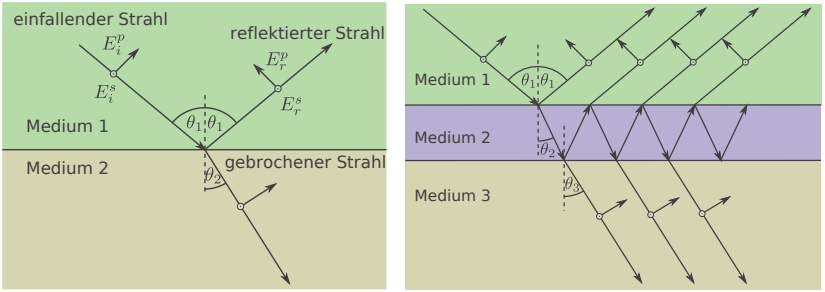
Abbildung 20.4: Orientierung von E_i^p, E_i^s und E_r^p, E_r^s bzgl. der Einfallsebene.

$$r_{12}^p = \frac{E_r^p}{E_i^p}$$
$$r_{12}^s = \frac{E_r^s}{E_i^s}$$

Hat man zwei Grenzflächen kann man durch Betrachtung von Mehrfachreflexionen über eine geometrische Reihe den Gesamtreflexionskoeffizienten r_{123}^p, r_{123}^s berechnen [6]:

$$\begin{aligned} r_{123}^p &= \frac{r_{12}^p + r_{23}^p \mathrm{e}^{-i2\beta}}{1 + r_{12}^p r_{23}^p \mathrm{e}^{-i2\beta}} \\ r_{123}^s &= \frac{r_{12}^s + r_{23}^s \mathrm{e}^{-i2\beta}}{1 + r_{12}^s r_{23}^s \mathrm{e}^{-i2\beta}} \\ \beta &= 2\pi \left(\frac{d}{\lambda}\right) n_2 \cos\theta_2 \end{aligned} \tag{20.1}$$

Dieser ist abhängig von r_{12}, r_{23}, der Wellenlänge λ und der Schichtdicke d. n_2 ist dabei der (komplexe) Brechungsindex der zweiten Schicht und θ_2 der Einfallswinkel auf die zweite Grenzschicht (siehe Abb. 20.5(b)). Bei isotropen Materialien gibt es keinen Energietransfer zwischen der p- und der s-polarisierten Welle. Der Polarisationszustand der reflektierten Welle kann auch bei einem Mehrschichtsystem durch separate Betrachtung der p- und der s-Polarisation berechnet werden. Bei anisotropen oder depolarisierenden Materialien muss auf den Müller-Formalismus zurückgegriffen werden.



(a) (b)

Abbildung 20.5: Reflexion an einer Grenzschicht (a) und Mehrfachreflexionen an zwei Grenzschichten (b).

3.3 Müller-Formalismus

Die allgemeinste Form zur Beschreibung des Polarisationszustands und von Zustandsänderungen des Lichts ist der Müller-Formalismus. Dieser beschreibt die physikalischen Vorgänge auch bei unvollständig polarisiertem Licht bzw. bei depolarisierenden Proben. Der Polarisationszustand wird dabei durch den Stokes-Parameter $\boldsymbol{S} \in \mathbb{R}^4$ beschrieben [7]:

$$S = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} I_{0^\circ} + I_{90^\circ} \\ I_{0^\circ} - I_{90^\circ} \\ I_{45^\circ} - I_{-45^\circ} \\ I_R - I_L \end{pmatrix}$$

Beim Stokes-Parameter werden Intensitäten gemessen, die sich bei Anwendung von Linearpolarisatoren unter $0^\circ, 45^\circ, 90^\circ, -45^\circ$ und links- bzw. rechtszirkularen Polarisatoren ergeben. Für beliebig polarisiertes Licht gilt:

$$S_0 = I_{0^\circ} + I_{90^\circ} = I_{45^\circ} + I_{-45^\circ} = I_L + I_R$$

Bei vollständig polarisiertem Licht gilt außerdem:

$$S_0^2 = S_1^2 + S_2^2 + S_3^2 \tag{20.2}$$

Eine Änderung des Polarisationszustands des einfallenden Lichtstrahls $\boldsymbol{S}_\mathrm{i}$ und des reflektierten Lichtstrahls $\boldsymbol{S}_\mathrm{r}$ durch die Probe lässt sich über die reelle Müller-Matrix $\mathbf{M} \in \mathbb{R}^{4\times4}$ beschreiben: $\boldsymbol{S}_\mathrm{r} = \mathbf{M}\boldsymbol{S}_\mathrm{i}$.

4 Ellipsometrische Messgrößen

Zwei elementare ellipsometrische Messgrößen sind Ψ und Δ. Diese sind definiert über die fundamentale Gleichung der Ellipsometrie:

$$\frac{r^p}{r^s} = \tan\Psi \mathrm{e}^{i\Delta} =: \rho$$

Ψ bringt die Änderung der Amplituden und Δ die Änderung der Phasen der p- und s-polarisierten Welle zueinander in Beziehung. Die Messwerte Ψ und Δ können mit einer einzelnen Messung erfasst werden, wenn E_i^s, E_i^p bekannt sind und E_r^s, E_r^p gemessen werden.

Im Retroreflexaufbau erfährt der Lichtstrahl mehrere Reflexionen. Es wird angenommen, dass der Retroreflektor auf den Polarisationszustand des Lichts keinen Einfluss hat. Dann ergibt sich der Gesamtreflexionskoeffizient in p-Polarisation zu $r'^p := (r^p)^2$ und in s-Polarisation zu $r'^s := (r^s)^2$. Bezüglich des Messwerts ρ beim klassischen Ellipsometer ergibt sich im Retroreflexaufbau $\rho' = r'^p/r'^s = \tan\Psi' \mathrm{e}^{i\Delta'} = \rho^2$. Für $\tan\Psi'$ und Δ' gilt dann: $\tan\Psi' = \tan^2\Psi$ und $\Delta' = 2\Delta$. Es gibt also eine Beziehung zwischen den Messwerten der klassischen Ellipsometrie und denen im Retroreflexaufbau, so dass Ψ und Δ zurückgerechnet werden können. Hierbei muss beachtet werden, dass der Phasenunterschied Δ verdoppelt wird und nur im Intervall $[0, \pi]$ bestimmt werden kann.

Aus Gl. (20.1) ist ersichtlich, dass der Reflexionskoeffizient und somit Ψ, Δ bzgl. der Schichtdicke eine Periodizität aufweisen. Diese beträgt $\lambda/(2\sqrt{n_2^2 - \sin^2\theta_1})$ [1]. Bei einem Einfallswinkel von $\theta_1 = 60°$ und der beim RRE eingesetzten He-Ne-Laserwellenlänge von 632,816 nm ergibt sich bei einer MgF_2-Beschichtung mit dem Brechungsindex $n_2 = 1,38$ die Periodizizät zu 294 nm. Die Ellipsometrie ist also ein sehr gutes Verfahren um kleine Schichtdickenvariationen zu messen. Bei größeren Variationen gibt es hingegen Mehrdeutigkeiten.

Beim RRE wird eine zirkular polarisierte Beleuchtung verwendet. Der Stokes-Parameter der eingestrahlten Strahlung ist $\boldsymbol{S}_i = (1, 0, 0, 1)$. Dieser Strahl wird mit der Müller-Matrix $\mathrm{M}_{\mathrm{Obj}}(\Psi', \Delta')$ der Probe multipliziert, die beide Reflexionen an der Probe beinhaltet. Nach doppelter Re-

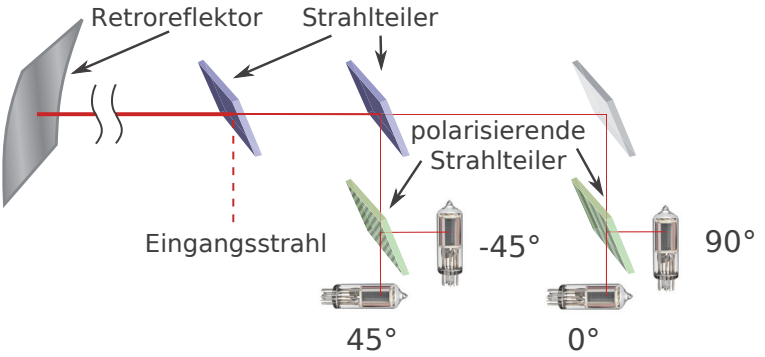


Abbildung 20.6: Darstellung des Strahlenverlaufs beim RRE und der unterschiedlichen Polarisationsmessungen unter $0°, 45°, 90°, -45°$ durch Photomultiplier.

flexion ergibt sich für den Stokes-Parameter:

$$\boldsymbol{S}_r = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} \mathbf{M}_{\text{Obj}} \cdot \boldsymbol{S}_i = \frac{|r'^s|^2}{2} \begin{pmatrix} 1 + \tan^2 \Psi' \\ 1 - \tan^2 \Psi' \\ 2 \tan \Psi' \sin \Delta' \\ 2 \tan \Psi' \cos \Delta' \end{pmatrix} \tag{20.3}$$

$\mathbf{M}_{\text{Obj}}$ ist definiert als [6]:

$$\mathbf{M}_{\text{Obj}} = \frac{|r'^s|^2}{2} \begin{pmatrix} 1 + \tan^2 \Psi' & 1 - \tan^2 \Psi' & 0 & 0 \\ 1 - \tan^2 \Psi' & 1 + \tan^2 \Psi' & 0 & 0 \\ 0 & 0 & 2 \tan \Psi' \cos \Delta' & 2 \tan \Psi' \sin \Delta' \\ 0 & 0 & -2 \tan \Psi' \sin \Delta' & 2 \tan \Psi' \cos \Delta' \end{pmatrix}$$

Beim jetzigen Sensoraufbau können S_0, S_1, S_2 direkt aus $I_{0°}$, $I_{90°}$, $I_{45°}$, $I_{-45°}$ gemessen werden. Der Aufbau des RRE zur Detektion des Polarisationszustands ist in Abb. 20.6 dargestellt. Bei nicht depolarisierenden Proben ist die Strahlung vollständig polarisiert und man kann über Gl. (20.2) auch $|S_3|$ aus S_0, S_1, S_2 berechnen. Da das Sensorkoordinatensystem nicht entsprechend der Einfallsebene ausgerichtet ist, muss noch eine Drehung des Sensorkoordinatensystems um den Winkel γ über die Drehmatrix $\mathbf{M}_{\text{Rot}}(\gamma)$ erfolgen. Wird der Stokes-Parameter $\boldsymbol{S}_{\text{Mess}} = (S_0^M, S_1^M, S_2^M, |S_3^M|)$ gemessen, so ist $\boldsymbol{S}_r = \mathbf{M}_{\text{Rot}}(-\gamma)\boldsymbol{S}_{Mess}$.

$\mathbf{M}_{\mathrm{Rot}}(\gamma)$ ist definiert als:

$$\mathbf{M}_{\mathrm{Rot}}(\gamma) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\gamma & -\sin 2\gamma & 0 \\ 0 & \sin 2\gamma & \cos 2\gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Somit gilt für $\boldsymbol{S}_r$:

$$\boldsymbol{S}_r = \begin{pmatrix} S_0^M \\ \cos 2\gamma S_1^M + \sin 2\gamma S_2^M \\ -\sin 2\gamma S_1^M + \cos 2\gamma S_2^M \\ |S_3^M| \end{pmatrix}$$

Aus $\boldsymbol{S}_r$ lässt sich Ψ' und Δ' im Intervall $[-\pi/2, \pi/2]$ nach Gl. (20.3) berechnen:

$$\tan \Psi' = \sqrt{\frac{S_0 - S_1}{S_0 + S_1}}$$

$$\sin \Delta' = \frac{S_2}{(S_0 + S_1) \tan \Psi'}$$

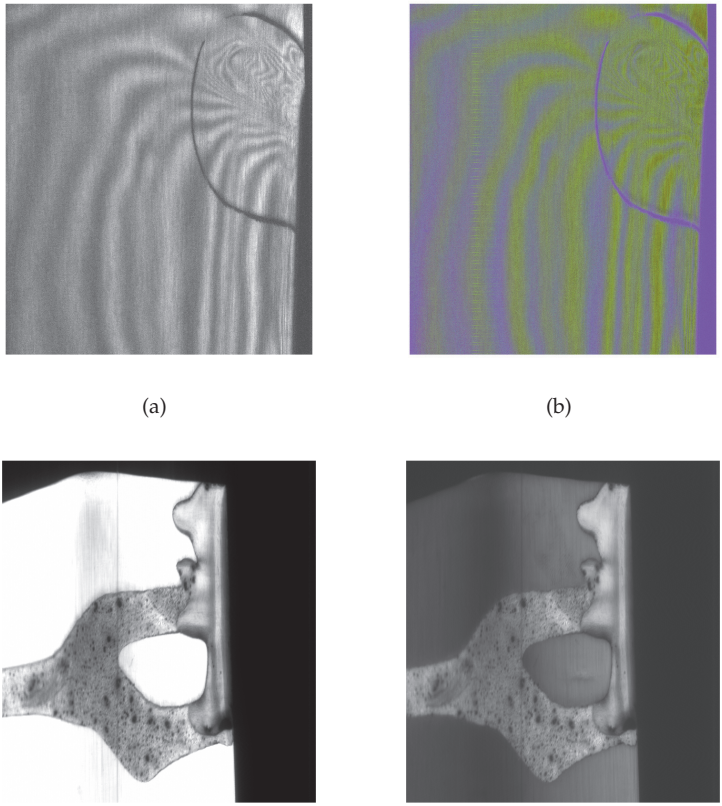
Aus $|S_3|$ lässt sich folgende Gleichung ableiten:

$$|\cos \Delta'| = \frac{|S_3|}{(S_0 + S_1) \tan \Psi'}$$

Dies bringt jedoch keinen Informationsgewinn über Δ, da folgende Beziehung gilt: $|\cos \Delta'| = \sqrt{1 - \sin^2 \Delta'}$

5 Polarisationsmessungen des RRE

Verschiedene Messproben wurden mit dem Ellipsometriescanner bereits untersucht. Die Verwendung eines Lasers als Lichtquelle in Kombination mit Photomultiplier zeigt sich hier als besonders vorteilhaft, da somit Messungen mit hohen Abtastraten auch an dunklen Proben durchgeführt werden können. In Abb. 20.7(a) ist der Kanal I_0 eines lackierten schwarzen Kunststoffteils abgebildet und in



(a) (b)

(c) (d)

Abbildung 20.7: Abbildung des I_{0°-Kanals (a) eines Kuststoffteils und der Hauptkomponentenanalyse über alle vier Intensitätskanäle (b). In (c) ist der I_{0°- und in (d) der I_{90°-Kanal eines Ölfilms auf einem Alu-Blech abgebildet.

Abb. 20.7(b) ist die Hauptkomponentenanalyse über die Intensitätswerte $I_{0^\circ}, I_{90^\circ}, I_{45^\circ}, I_{-45^\circ}$ in Falschfarbendarstellung dargestellt. Es sind Modulationsstreifen sichtbar, die durch Schichtdickenvariationen der Lackbeschichtung und Neigungsänderungen entsprechend Gl. (20.1) entstehen. Der markierte Defekt sind Inhomogenitäten in der Schichtdicke. Diese treten bevorzugt am Rand einer Oberfläche bei ändernder Oberflächenspannung auf. In Abb. 20.7(c) und 20.7(d) sind I_{0° bzw. I_{90° eines Ölfilms auf einer Alu-Oberfläche abgebildet. Während im Ölfilm Modulationsstreifen sichtbar sind, zeigt das Alu-Blech eine deutlich höhere Intensität bei der s-polarisierten Strahlung gegenüber der p-polarisierten, was mit theoretischen Überlegungen übereinstimmt. Eine Depolarisation durch den Retroreflektor oder ein Energietransfer zwischen p- und s-Polarisation findet daher nicht statt. Andere Messungen haben ergeben, dass der verwendete Retroreflektor bei starker Variation des Einfallswinkels in den Polarisationskanälen nahezu die selbe Intensität erhält, so dass die Müller-Matrix als Einheitsmatrix angenommen werden kann.

6 Zusammenfassung und Ausblick

In diesem Artikel wurde gezeigt, dass bildgebende ellipsometrische Messungen an gekrümmten Oberflächen möglich sind. Zudem wurde dargestellt, wie aus den erhaltenen Bilddaten klassische Messwerte berechnet werden können. Eine Berechnung der Schichtdicke ist zu diesem Zeitpunkt noch nicht möglich, da sowohl der Einfallswinkel als auch die Drehung des Sensorkoordinatensystems bzgl. der Einfallsebene bekannt sein muss. Liegt ein CAD-Modell der Probe vor, wie das in der Automobilindustrie häufig der Fall ist, lässt sich beides daraus berechnen. Sind außerdem noch Sollwerte für die Schichtdicken bekannt, kann man Bilddaten offline simulieren und sie mit den gemessenen Bildern vergleichen. Eine Sensordatenfusion mit einem zusätzlichen Messsystem zur Geometrievermessung ist auch denkbar, wodurch dann kein A-priori-Wissen über die Geometrie der zu prüfenden Oberfläche notwendig ist. Abhängig vom Anwendungsfall sind auch Modifikationen des jetzigen Sensoraufbaus vorstellbar um den Stokes-Parameter vollständig zu bestimmen. Damit können sowohl depolarisierende Oberflächen vermessen als auch Δ im Inter-

vall $[0, \pi]$ bestimmt werden. Mehrdeutigkeiten bei der Schichtdickenbestimmung werden dadurch verringert. Auch Bildverarbeitungsalgorithmen können zur Auflösung von Mehrdeutigkeiten implementiert werden, wenn über die Schichtdicke Glattheitsannahmen getroffen werden. Durch Variation des Einfallswinkels von benachbarten Pixeln können über Datenfusion Mehrdeutigkeiten aufgelöst werden. Auf eine ähnliche Weise werden bei der spektroskopischen Ellipsometrie Mehrdeutigkeiten durch Variation der Wellenlänge aufgelöst. Ein weiterer geplanter Arbeitspunkt ist die Vermessung der Müller-Matrix des Retroreflektors. Diese kann aufgrund der Rückreflexion nicht mit herkömmlichen Müller-Matrix-Ellipsometern erfasst werden, sondern muss mit dem RRE-Ellipsometer vermessen werden.

Literatur

1. H. G. Tompkins und W. A. McGahan, *Spectroscopic ellipsometry and reflectometry: A user's guide*. New York: Wiley, 1999.
2. L. Asinovski, D. Beaglehole und M. T. Clarkson, „Imaging ellipsometry: quantitative analysis“, *physica status solidi (a)*, Vol. 205, Nr. 4, S. 764–771, 2008.
3. U. Neuschaefer-Rube, Hrsg., *Optische Oberflächenmesstechnik für Topografie und Material*, Ser. Fortschritt-Berichte VDI Reihe 8, Meß-, Steuerungs- und Regelungstechnik. Düsseldorf: VDI-Verl., 2002, Vol. 953.
4. H. Fu, T. Goodman, S. Sugaya, J. K. Erwin und M. Mansuripur, „Retroreflecting ellipsometer for measuring the birefringence of optical disk substrates“, *Applied optics*, Vol. 34, Nr. 1, S. 31–39, 1995.
5. Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V., „Vorrichtung und verfahren zur optischen charakterisierung von materialien / apparatus and method for optically characterizing materials“, Patente EP11760701.0 (09.09.2011), DE 10 2010 046 438.4-52 (29.03.2012), WO 2012/038036 A1 (29.03.2012), US-2013-0222803-A1 (29.08.2013).
6. J. Beyerer, F. Puente León und C. Frese, *Automatische Sichtprüfung: Grundlagen, Methoden und Praxis der Bildgewinnung und Bildauswertung*, Ser. SpringerLink : Bücher. Berlin and Heidelberg: Springer, 2012.
7. H. Fujiwara, *Spectroscopic ellipsometry: Principles and applications*. Chichester and England and Hoboken and NJ: John Wiley & Sons, 2007.

Verbesserung von Positionsbestimmungen mittels holografischer Mehrpunktgenerierung

Tobias Haist, Marc Gronle, Thomas Arnold, Duc Anh Bui
und Wolfgang Osten

Institut für Technische Optik, Universität Stuttgart
Pfaffenwaldring 9, 70569 Stuttgart

Zusammenfassung Vorgeschlagen wird eine Methodik zur Verbesserung der Messunsicherheit für bildverarbeitungsbasierte Positionsbestimmungen. Hierzu wird vor dem abbildenden Objektiv ein Hologramm als Vervielfältigungselement eingesetzt. Ein einzelner heller Objektpunkt wird so in der Ebene des Bildsensors in N Punkte aufgespalten. Statistische Fehler sowie Diskretisierungsfehler vermindern sich bei der Mittelwertbildung der Positionsbestimmungen der N Punkte um $\sqrt{N}$. Statistische Messunsicherheiten deutlich unterhalb 1/100 Pixel lassen sich auf diese Weise mit preisgünstigen Standardsensoren erzielen.

1 Einleitung

Die genaue bildbasierte Bestimmung von Positionen ist sowohl in der zwei- als auch der dreidimensional messenden Bildverarbeitung von zentraler Bedeutung und Basis verschiedener Messmethodiken zur Ermittlung geometrischer Größen bzw. Positionen. Beispiele sind konventionelle 2D Anwendungen, Shack-Hartman Sensoren und triangulierende Sensoren. Hohe Genauigkeiten im Bereich von 1/10 bis zu 1/100 Pixel sind unter idealen Bedingungen beim Stand der Technik mit Subpixel-Algorithmen bei sehr gut kalibrierten Systemen erreichbar [1].

Im Folgenden soll eine Methodik vorgestellt werden, die in der Lage ist, diese Genauigkeit weiter zu steigern. Hierzu werden erste experimentelle Ergebnisse vorgestellt. Als Basis für die Darstellung soll zunächst kurz auf die hier zentrale Größe der Positionsauflösung und die Messunsicherheit von Positionsbestimmungen eingegangen werden.

Der Begriff „Auflösung" wird in der Optik anwendungsabhängig sehr unterschiedlich definiert bzw. benutzt. Unterschieden werden muss zwischen Auflösung im Sinne

- der Unterscheidung bzw. Trennung eng benachbarter Strukturen (z.B. Zweipunkt-Trennung),
- der Bestimmung der Position eines Punkts oder Linie (Positionsauflösung),
- der Auflösung einer axialen Messung (Höhen- bzw. „3D-Messung") und
- der reinen Auflösung eines Bildsensors als Anzahl der verfügbaren Pixel.

Die unterschiedlichen Auflösungen hängen zwar teilweise miteinander zusammen, sind aber in keinem Fall durcheinander ersetzbar. In der Bildverarbeitung wird der Begriff „Auflösung" ohne nähere Kennzeichnung meist im Sinne der Unterscheidung eng benachbarter Strukturen benutzt und dann aber oft (und fälschlicherweise) mit der Positionsauflösung gleichgesetzt.

Die Auflösung im Sinne der Trennung von Strukturen kann unterschiedlich spezifiziert und berechnet werden. Für benachbarte isolierte Strukturen (Punkte) ergibt sich das theoretische Maximum aufgrund der Beugungsbegrenzung für ein ideales System durch die Auflösung nach Rayleigh im Bildraum $r'_R = 0.61\,\lambda\,K$ mit der Blendenzahl K und der Wellenlänge λ oder durch vergleichbare Kennzahlen (z.B. Auflösung nach Sparrow). Für periodische Strukturen ist die Auflösung nach Abbe im Bildraum $r'_A = 0.5\,\lambda\,K$ oder aber eine Angabe mittels der Modulationstransferfunktion (MTF) entscheidend [2]. Für nicht-ideale Systeme sind entsprechend ebenfalls die MTF (für periodische Strukturen) oder aber mittlere Spotradien der Punktbildfunktion (Einzelstrukturen) verwendbar.

Für die Messung einer geometrischen Größe (z.B. Bohrlochdurchmesser) ist allerdings nicht die durch diese Auflösungskennzahlen letztlich erfasste Ausdehnung des Punktbilds von primärem Belang. Wesentlich sind hier andere Faktoren, die sich aus der Abfolge der Arbeitsschritte zur Messung der geometrischen Größe ergeben. In der Regel wird zunächst das Bild hinsichtlich seiner Verzeichnung korrigiert (systematische Fehler), dann werden die relevanten Kanten oder Struktu-

ren subpixelgenau detektiert [3]. Entlang der Kanten kann schließlich lateral gemittelt werden, um statistische Fehler zu unterdrücken. Für das Beispiel eines Bohrlochs wird man beispielsweise alle oder zumindest viele Bildpunkte, die den Bohrlochradius definieren nutzen und so mitteln. Bei N Punkten ergibt sich damit eine Verbesserung um den Faktor $\sqrt{N}$ für unkorrelierte Fehler. Diese Mittelung kann explizit erfolgen oder aber auch implizit im Algorithmus enthalten sein (z.B. Hough-Transformation).

Die Genauigkeit der Kantendetektion ist wieder von einer Vielzahl von Parametern abhängig, letztlich ist aber eine gewisse Unschärfe der Kante vorteilhaft bzw. sogar notwendig, um hohe Subpixelgenauigkeiten zu erzielen. Das ideale Maß an Unschärfe hängt dabei in komplexer Weise von den verwendeten Algorithmen, dem Rauschen, den Objekten und weiteren Parametern ab.

Die Grundidee des hier vorgeschlagenen Verfahrens (vgl. Abschnitt 4) ist es, den wesentlichen Teil der statistischen Fehler der einzelnen Positionsdetektion durch eine Vervielfältigung der Bildinformation *vor* der Bildaufnahme zu verringern. Wir betrachten in dieser Arbeit den einfachsten Fall, nämlich die Detektion der Position isolierter Punkte.

2 Positionsbestimmung idealer Punkte

Die Genauigkeit, mit der die Position des (beugungsbegrenzten) Bildes eines idealen Punktes gemessen werden kann, hängt einerseits von der Ausdehnung der Lichtverteilung (entsprechend z.B. der Auflösung nach Rayleigh) und andererseits von der Anzahl der Photonen, die erfasst werden, ab. Bei einer gegebenen Standardabweichung s einer Lichtverteilung und der Detektion von M Photonen kann durch Schwerpunktsberechnung bei Rauschfreiheit der Detektion eine Positionsbestimmung mit der Genauigkeit (Standardabweichung) von $s' = s/\sqrt{M}$ erzielt werden [4]. Dabei geht man von einer beliebigen Anzahl von Pixeln mit einer beliebig hohen Quanten-Well Kapazität (und Quanteneffizienz 1) aus. Die einfache Schwerpunktsbestimmung liefert für diesen Fall die ideale Positionsschätzung.

In der Praxis ergeben sich natürlich gegenüber dieser Idealvorstellung zahlreiche Störungs- und Rauscheinflüsse (u.a. Diskretisierung, Photonenrauschen, verschiedene elektronische Rauscheinflüsse,

Quantisierung) [5]. Einen Überblick über in der Praxis erzielbare Genauigkeiten findet man in [1, 6–8]. Üblicherweise lassen sich RMS-Abweichungen von der exakten Position von bis zu 1/100 Pixel erreichen.

Die erzielbare Positionsauflösung wird von diesen Einflüssen bestimmt. Dominant sind dabei aber meist statistische Fehler sowie die Diskretisierung. Beide Fehler lassen sich um den Faktor $\sqrt{N}$ verringern, indem die subpixelgenaue Position von N Punkten bestimmt und dann der arithmetische Mittelwert der N Positionsbestimmungen verwendet wird. Auch hierbei kann die Mittelung explizit oder aber auch implizit (z.B. durch Korrelationsverfahren) durchgeführt werden.

Genutzt wird dies seit langer Zeit beispielsweise in der Photogrammetrie bei der Verwendung von ausgedehnten Targets, die auf einem zu vermessenden Objekt befestigt werden. Dies führt aber zu drei wesentlichen Nachteilen. 1) Die Markierung des Objekts ist aufwändig und fehleranfällig, 2) die Anzahl der Messpunkte auf dem Objekt wird stark begrenzt und hohe laterale Auflösungen sind prinzipbedingt nicht möglich und 3) es ergeben sich starke Einschränkungen hinsichtlich der Genauigkeit bei gekrümmten und geneigten Oberflächen.

3 Holografische Vervielfachung

Für viele Anwendungen praktikabler ist die Vervielfältigung der Positionsinformation nicht auf dem Objekt sondern erst vor der Bildebene des Bildsensors, denn letztlich begrenzt dieser aufgrund Diskretisierung, Rauschen, Quantisierung und weiterer Fehlereinflüsse die erreichbare Positionsgenauigkeit. Dies wird durch das Hologramm (Abb. 21.1) erreicht. Diese Methodik funktioniert bei einem einfachen Hologramm zunächst nur bei quasi-monochromatischem (aber durchaus inkohärentem) Licht und natürlich nur wenn genügend Licht vorhanden ist, um die erhöhte Anzahl der bestrahlten Pixel ausreichend zu belichten.

Eine Simulation des Vorgangs zeigt die erwartete Verbesserung der erzielbaren statistischen Messunsicherheit mit der Wurzel der Anzahl der Punkte. Details hierzu finden sich in [9].

Abbildung 21.1: Prinzip der Bildvervielfachung durch ein computergeneriertes Hologramm (CGH) am Beispiel der Abbildung eines Punktes. Jeder objektseitige Punkt führt kameraseitig zu N Punkten, deren Position ermittelt wird. Eine Verschiebung des Objektpunkts führt im Bildraum zu einer simultanen Verschiebung aller N Punkte.

4 Experimentelle Ergebnisse

Erste experimentelle Ergebnisse wurden mit einem Laser als Lichtquelle für einen sehr kleinen Bereich (ca. 0,2) Pixel kameraseitig gewonnen (vgl. Abb. 21.2). Dargestellt sind die ermittelten Positionen von Einzelpunkten sowie für die gemittelten Punkte bei einer linearen Subpixelverschiebung des Spots. Für die Positionsbestimmung der einzelnen Positionen wurde der Schwerpunkt in einem Fenster um das Spotmaximum auswertet. Dabei wurden alle Intensitäten unterhalb einer rauschabhängigen Schelle (3 x Rauschamplitude) nicht berücksichtigt. Das Hologramm ist ein einfaches binäres Element (Photoresist auf Glas, optimiert mit einem direkten Suchverfahren).

Gegenüber der idealen Geraden ergibt sich eine RMS-Abweichung von 0,0028 Pixeln im Vergleich zu 0,010 Pixel für Einzelspots für die Mittelung von 14 Einzelposition. Eine genaue Beschreibung des verwendeten Messaufbaus findet sich in [9].

Bei den extrem hohen Messunsicherheiten im Bereich kleiner 1/100 Pixel müssen hier eine Vielzahl von Fehlerquellen in Betracht gezogen werden. Wesentlich sind vor allem die konsequente Abschirmung gegenüber Luftturbulenzen und die Vermeidung jeglicher mechanischer Vibrationen. Kameraseitig sind natürlich alle relevanten Rauschparameter (Ausleserauschen, thermisches Rauschen, Fixed-Pattern Rauschen [10] und räumliche Variation der Empfindlichkeit) und Pixelin-

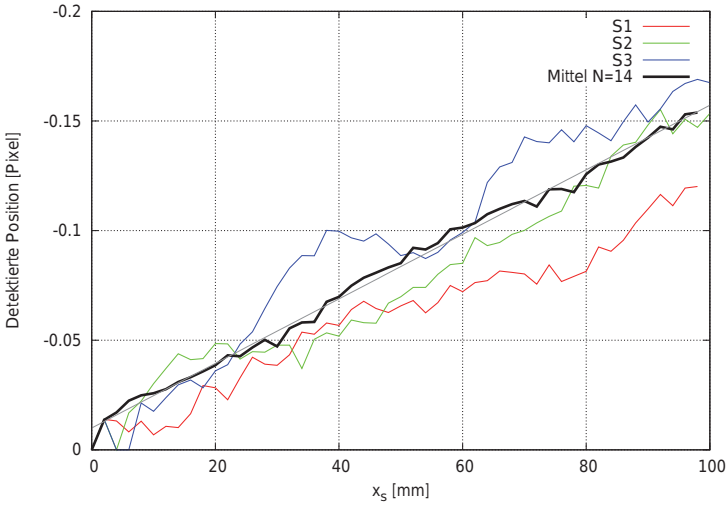


Abbildung 21.2: Verschiebungsbestimmung für einzelne Punkte sowie 14 Punkte gemittelt bei 20 ms Belichtungszeit mit Sony ICX655AQA Sensor.

stabilitäten (Jitter) von Belang.

Zu beachten ist aber, dass für eine Messung über mehrere Pixel die Diskretisierungseffekte verstärkt auftreten können und so die Genauigkeit reduzieren. Details hierzu hängen natürlich von der exakten Pixelgeometrie (bzw. der Kombination aus Mikrolinsen und Sensor) ab. Die Mittelung über N Punkte verbessert auch hier deutlich das Ergebnis denn die Diskretisierungsfehler der N Punkte sind nicht korreliert und werden daher durch die Mittelung reduziert.

Das Verhalten für größere Verschiebungen und bei Beleuchtung mit LEDs wurde in einem erweiterten Versuchsaufbau mit interferometrischer Positionskontrolle durchgeführt.

Abb. 21.3 zeigt ein Beispiel über einen Bereich von 2,2 Pixeln, der mit einem Sony ICX655AQA Sensor (SVS Vistek eco655 Kamera , monochrom, Gigabit Ethernet (GigE) interface, 2448 x 2050 Pixel, 3,45 μm Pixel, 60 MHz Auslesefrequenz, Ausleserauschen kleiner 10 Elektronen, Quantum-well Kapazität 8000 Elektronen, in 8 bit Mode bei einer Belichtungszeit von 20 ms genutzt) aufgenommen wurde. Zur Bestim-

mung der Referenzposition wurde ein SIOS Dreistrahlinterferometer (SP 2000 TR) verwendet. Als Lichtquelle diente eine rote LED (Osram LRW 5SN-JYKY1, 24,8 cd, $\lambda = 632$ nm) in Kombination mit einem Interferenzfilter (Thorlabs FL05632-1, $\lambda = 632{,}8$ nm, FWHM = 6 nm (eigene Messung)) zur Vermeidung chromatischer Aberrationen und eine schrittmotorgesteuerte Verfahrung der LED mit einem 1:50 Abbildungssystem (Abbildungsobjektiv Lensagon CMFA1520ND, Brennweite, 15 mm, eingesetzt bei Blendenzahl 2). Luftfluktuationen wurden durch zylindrische Röhren zur Einfassung des optischen Strahlengangs reduziert.

Aufgrund von Restfehlern (Vibrationen, Luft, Jitter) wurden jeweils 30 Bilder bei einer Position aufgenommen und ausgewertet. Dabei erhöht sich weiterhin die effektiv wirksame Quantum-well Kapazität (Faktor 30) sowie das thermische Rauschen und das Ausleserauschen (Faktor $\sqrt{30}$). Die Standardabweichung der Abweichung von einer Ausgleichsgeraden liegt bei 0,0042 Pixel (entsprechend 14,5 nm). Für den Einzelpunkt erhält man 0,0067 Pixel. Die Verbesserung durch die Mittelung über 16 Punkte erzielt in diesem Fall zwar eine deutliche Verbesserung (37%), entspricht aber nicht dem erwarteten Faktor 4. Grund hierfür ist die bereits sehr gute Unterdrückung statistischer Fehler (starke zeitliche Mittelung und sehr gute Abschirmung von Luftturbulenzen), so dass verbleibende, teilweise systematische Fehler (vermutlich Diskretisierungsfehler und Jitter) einen größeren Einfluss am Gesamtfehler haben.

Ein noch besseres Ergebnis wäre vermutlich erzielbar, wenn zunächst jede Einzelposition gemäß dem erwarteten systematischen Verhalten korrigiert werden würde bevor gemittelt wird. Zukünftige Arbeiten in diese Richtung sollten es ermöglichen, die extrem geringen Messunsicherheiten so auch für ausgedehnte Positionen auf dem Bildaufnehmer weiter zu verringern.

5 Diskussion

Da es sich bei den dargestellten Ergebnisse um Einzelpunktergebnisse handelt kann bei der Antastung von ausgedehnten Kanten unter idealen Bedingungen noch mit einer weiteren Genauigkeitssteigerung (Mittelung entlang der Kante) gerechnet werden. Hierbei ist allerdings

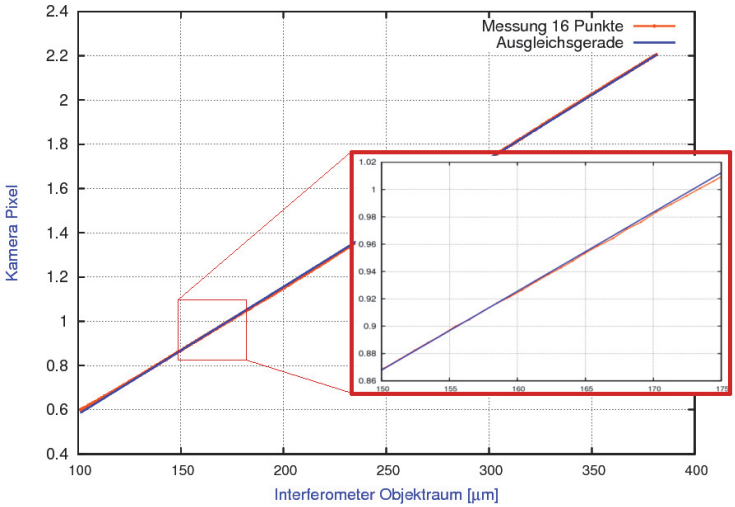


Abbildung 21.3: Messung unter interferometrischen Kontrolle über einen Verschiebungsbereich von 2,2 Pixeln. Die Standardabweichung der ermittelten von der wahren Position beträgt 0,0042 Pixel.

zu beachten, dass sich Einschränkungen an das Objekt ergeben, um eine sensorseitige deutliche Überlappung von Kanteninformation zu vermeiden.

Aufgrund der holografischen Vervielfältigung der Bildinformation ist eine ausreichend hohe zeitliche Kohärenz erforderlich. Die Anforderungen sind allerdings für typische Abbildungsgeometrien gering. Entscheidender ist die (radiale) Verbreiterung der abgebildeten Bildpunkte durch chromatische Aberration, so dass eine breitbandige Beleuchtung zu Problemen führt. Weiterhin muss genügend Licht zur Verfügung stehen, um den Bildsensor an den N Positionen ausreichend zu belichten.

Generell soll nochmals betont werden, dass es sich bei den dargestellten Ergebnissen, die sensorseitig im Bereich tausendstel Pixel bzw. Nanometer liegen, um statistische Messunsicherheiten handelt. Systematische Fehler müssen also in einer praktischen Anwendung unterdrückt bzw. kalibriert werden. Insbesondere macht ein Einsatz der Methodik nur in Kombination mit einer sehr guten Kamerakalibrie-

rung (oder symmetrischen Abbildungssystemen) für global messende Aufgabenstellungen Sinn. Die Verschiebungsmessung über kleinere Bereiche kann auch mit geringeren Anforderungen hinsichtlich der Kalibrierung vorgenommen werden.

Wir danken der Deutschen Forschungsgemeinschaft für die finanzielle Unterstützung im Projekt DFG-OS 111/42-1.

Literatur

1. M. R. Shortis und T. A. Clarke, „Practical Testing of the Precision and Accuracy of Target Image Centring Algorithms“, *Proc. SPIE*, Vol. 2598, S. 65–76, 1995.
2. A. Hornberg, *Handbook of Machine Vision*, 1. Aufl. WILEY-VCH, 2006.
3. C. Demant, B. Streicher-Abel und A. Springhoff, *Industrielle Bildverarbeitung*. Springer, 2011.
4. S. Thomas, „Optimized centroid computing in a shack-hartmann sensor“, *Proc. SPIE*, Vol. 5490, S. 1238–1246, 2004.
5. D. R. Neal, J. Copland und D. A. Neal, „Shack-hartmann wavefront sensor precision and accuracy“, *Proc. SPIE*, Vol. 4779, S. 148–160, 2002.
6. S. Wang, B. Yan, M. Dong, J. Wang und P. Sun, „An improved centroid location algorithm for infrared led feature points“, *Proc. SPIE*, Vol. 8916, S. 891 619–891 619–9, 2013.
7. L. Bo, D. Mingli, J. Wang und Y. Bixi, „Sub-pixel location of center of target based on Zernike moment“, *Proc. SPIE*, Vol. 7544, S. 75 443A–75 443A–6, 2010.
8. J. Trinder, J. Jansa und Y. Huang, „An assessment of the precision and accuracy of methods of digital target location“, *Journal of Photogrammetry and Remote Sensing*, Vol. 50, Nr. 2, S. 12 – 20, 1995.
9. T. Haist, S. Dong, T. Arnold, M. Gronle und W. Osten, „Multi-image position detection“, *Opt. Express*, Vol. 22, Nr. 12, S. 14 450–14 463, 2014.
10. C. Li, M. Xia, Z. Liu, D. Li und L. Xuan, „Optimization for high precision Shack-Hartmann wavefront sensor“, *Optics Communications*, Vol. 282, Nr. 22, S. 4333–4338, 2009.

Positionierungsverifikation komplexer Großbauteile in der Roboterzelle zur Erweiterung eines Prozessführungssystems für die Automatisierung von MRO-Prozessen

Jan Philipp Steinbach[1], Tobias Ernst[1], Alexander Fay[1]
und Florian Hartung[2]

[1] Helmut-Schmidt-Universität,
Institut für Automatisierungstechnik,
Holstenhofweg 85, 22043 Hamburg
[2] Lufthansa Technik AG,
Weg beim Jäger 193, 22335 Hamburg

Zusammenfassung In diesem Paper wird ein Verfahren für die Verifikation der korrekten Positionierung eines Bauteils eines großen Bauteilspektrums vorgestellt, welches auf einem angepassten Hintergrundsubtraktionsverfahrens basiert. Aufgrund seiner einfachen Handhabbarkeit eignet sich das Verfahren zur kostengünstigen Erweiterung von Prozessführungssystemen in der Produktionsumgebung. Die Evaluation der Ergebnisse erfolgt in der realen Industrieumgebung.

1 Motivation

Die Wartung, Reparatur und Überholung von Flugzeugkomponenten, die zwar wiederkehrend, aber lediglich in kleinen Stückzahlen und Losgrößen sowie einem weitreichenden Bauteilspektrum auftreten, erfordert flexible Produktionsprozesse. So ist der Maintenance, Repair and Overhaul (MRO) Markt durch die Bearbeitung kleiner Stückzahlen hochpreisiger und hochkomplexer Produkte mit langen Produktlebenszyklen gekennzeichnet, deren Handhabung und Bearbeitung durch eine Vielzahl restriktiver Bestimmungen reglementiert ist [1]. Der für die Jahre 2007 bis 2017 von der Branche prognostizierte Anstieg der weltweit eingesetzten Flugzeuge um 50% führt einerseits zu starken

Wachstumsprognosen für die MRO-Branche selbst, erfordert andererseits auch von dieser, bestehende MRO-Prozesse zu automatisieren und effizienter sowie wirtschaftlich rentabler zu gestalten [1]. Neben der Anwendung von Konzepten aus dem Bereich der Lean Automation auf MRO-Prozesse [2] müssen dazu auch die innerhalb der job shops noch überwiegend manuell durchgeführten Bearbeitungsprozesse durch den Einsatz von Industrierobotern und hochtechnisierten Anlagen automatisiert werden. Dazu gilt es, durch die Kombination flexibler und reproduzierbar einstellbarer Spannsysteme mit dem Werker als flexibles Prozesselement und geeigneten hochtechnisierten Anlagen und Robotern den bisherigen Automatisierungsgrad eines MRO-Prozesses zu erhöhen. Die Akzeptanz der Integration neuer Produktionstechnologien in derartige Prozesse, die über Jahrzehnte handwerklich orientiert waren, kann durch die parallele Integration eines Prozessführungssystems deutlich gesteigert werden. Besonderes Augenmerk ist bei einer derartigen Teilautomatisierung eines MRO-Prozesses auf die Schnittstellen zwischen Mensch, Prozessführung und Anlage zu legen, um die Fehler und deren Auswirkungen aufgrund menschlichen Versagens zu minimieren. Insbesondere die robotergestütze Bearbeitung erfordert die exakte Positionierung des Bauteils gegenüber dem Roboterkoordinatensystem. Der Herausforderung der reproduzierbaren Positionierung eines großen Bauteilspektrums in der Roboterzelle lässt sich mit flexiblen und reproduzierbar konfigurierbaren Spannsystemen begegnen. Dennoch bleibt, trotz Verwendung einer Prozessführung, der die jeweiligen Spannpositionen für ein zu bearbeitendes Bauteil zu entnehmen sind, ein hohes Risiko der fehlerhaften Positionierung eines Bauteils. Die Auswirkungen reichen von der Bearbeitung in mangelhafter Qualität bis zur Kollision von Roboter und Bauteil. Daher sollte im hier exemplarisch im Mittelpunkt stehenden Projekt die Prozessführung für einen berührungslosen Bearbeitungsprozess von Großbauteilen mittels eines 8-achsigen Portalroboters um ein System zur Verifikation der korrekten Positionierung des zu bearbeitende Bauteils erweitert werden. Dazu wurde das im Folgenden vorgestellte Verfahren zur Hintergrundsubtraktion entwickelt, das eine kostengünstige Erweiterung des Prozessführungssystems und die benutzerfreundliche Handhabung für das Bedienpersonal ermöglicht. Primärer Fokus liegt dabei auf der Positionierungsverifikation der Bauteile, die auf einer Spannvorrichtung reproduzierbar fixiert werden.

2 Verfahren zur Hintergrundsubtraktion

Die Hintergrundsubtraktion ist eine bekannte Methode zur Erkennung von Vordergrundobjekten in Bilddaten und findet insbesondere im Bereich der Video- und Verkehrsüberwachung häufig Anwendung. Seit den frühen 1990er Jahren ist die Hintergrundsubtraktion Schwerpunkt vieler Forschungstätigkeiten und Veröffentlichungen [3], die in der Vergangenheit zu einer Vielzahl an entwickelten Verfahren und Algorithmen geführt haben. Die klassischen Ansätze zur Hintergrundmodellierung basieren auf der Modellierung eines Hintergrundpixels anhand von Wahrscheinlichkeitsdichtefunktionen oder statistischen Parametern wie dem Mittelwert oder der Standardabweichung. Andere vorgestellte Methoden zur Hintergrundsubtraktion modellieren den Hintergrund anhand von Hintergrundmustern [4, 5].

Die Vielzahl an in der Vergangenheit veröffentlichten Ansätzen und Verfahren führte auch zur Veröffentlichung mehrerer Vergleiche existierender Hintergrundsubtraktionsalgorithmen. Dabei erfolgen Evaluationen häufig im Kontext konkreter Anwendungen aus dem Bereich der Videoüberwachung [6–8]. In den letzten Jahren wurden Evaluationen jedoch auch verstärkt anhand großer systematisch erstellter reproduzierbarer Datensätze durchgeführt [3, 4, 9], die grundsätzliche Anforderungen an Videoüberwachungsalgorithmen repräsentieren.

Der Ansatz dieser Veröffentlichung, eine Prozessführung um ein System zur Verifikation der korrekten Positionierung komplexer Großbauteile in der Roboterzelle zu erweitern, basiert auf der Kombination von Verfahren zur Hintergrundsubtraktion aus dem Anwendungskontext der Videoüberwachung. Die Algorithmen, die zur Adaption des Verfahrens an die Applikation kombiniert wurden, werden im folgenden Abschnitt im Detail erläutert.

3 Positionserkennung durch Hintergrundsubtraktion

Die Adaption und Erweiterung eines Verfahren zur Hintergrundsubtraktion aus dem Kontext der Videoüberwachung ermöglicht eine robuste Trennung des Vordergrundes, bestehend aus Bauteilaufnahme und Bauteil, von dem höchst suboptimalen Hintergrund der Roboterzelle. Um die Robustheit des Verfahren hinsichtlich des großen Spek-

trums an zu erkennenden Bauteilen zu steigern, setzt sich das folgend vorgestellte Bildverarbeitungssystem aus zwei oder mehr Kameras zusammen, die an Seitenfenstern und ggf. an der Decke einer großen Roboterzelle ($5\,\mathrm{m} \times 5\,\mathrm{m} \times 5\,\mathrm{m}$) fixiert sind. Abbildung 22.1 zeigt den schematischen Aufbau des Systems sowie eine reale Aufnahme der Roboterzelle.

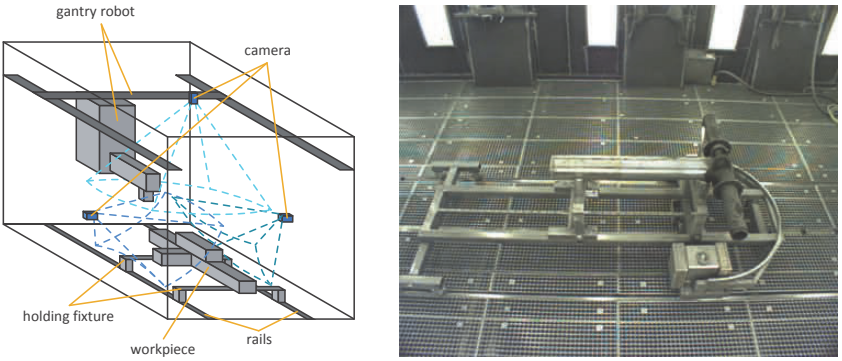


Abbildung 22.1: Skizze der Roboterzelle (a) und reale Aufnahme der Roboterzelle mit exemplarisch positioniertem Bauteil (b).

Die Verifikation der korrekten Positionierung eines Bauteils erfolgt für jede Kamera mittels Template Matching der zu prüfenden Aufnahme und eines jeweiligen Referenzbildes. Auf beiden Bildern wird dazu zuvor das folgend vorgestellte Verfahren zur Hintergrundsubtraktion angewendet. Dies ermöglicht die robuste Segmentierung des für die Prüfung relevanten Vordergrundes, auch unter den für die Bildverarbeitung suboptimalen Bedingungen hinsichtlich Hintergrund, Beleuchtung und Reflexionen.

3.1 Grundlegender Ansatz der Hintergrundsubtraktion

Grundlegender Schritt für die Verifikation der korrekten Bauteilpositionierung ist die Trennung von Vorder- und Hintergrund in den aufgenommenen 2-D Kamerabildern. Dabei ist zu berücksichtigen, dass die

nur im Zuge der Verifikation aufgenommenen Bilddaten sowie die statische Szenerie innerhalb einer Produktionszelle kaum zu dynamischem Verhalten führen, aus dem sich viele Informationen gewinnen lassen. Daher wurde als grundlegendes Verfahren zur Hintergrundsubtraktion das von Jabri [10] vorgestellte Verfahren gewählt. Dieser Algorithmus verwendet neben den Farbinformationen, auf denen der Großteil aller Hintergrundsubtraktionsverfahren basiert, auch Kanteninformationen zur Hintergrundmodellierung und zur Vordergrundsegmentierung. Zusätzlich wurde das Verfahren nach Jabri mit dem musterbasierten Ansatz der nicht-statistisch basierten Verfahren von Barnich [11] und Hofmann [5] kombiniert. Dies ermöglicht die Updatefähigkeit des Hintergrundmodells auch bei geringer Dynamik.

Modellierung des Hintergrundes

Die Modellierung des Hintergrundes erfolgt anhand eines Farb- sowie eines Kantenmodells, das aus vertikalen und horizontalen Kanten des Hintergrundes besteht. Für die Erstellung des Hintergrundmodells werden zunächst n Hintergrundbilder $I_{bck,i}$ aufgezeichnet, deren Farb- und Kanteninformationen in Form von Musterklassen gespeichert werden. Für jedes Muster existieren folglich n Datensätze. Aus diesen Datensätzen aller Muster erfolgt gemäß Jabri [10] die Berechnung des Farbmodells in Form der mittleren Farbwerte $\overline{F}_{k,P}$ und der Standardabweichung $\sigma_{k,P}$ für jeden Farbkanal k und jedes Pixel P_{k,x_P,y_P}. Das Kantenmodell setzt sich aus den horizontalen und vertikalen Kantenbildern jedes Farbkanals des Hintergrundes $I_{h,bck,i}$ bzw. $I_{b,bck,i}$ zusammen. Für diese werden ebenfalls die mittleren Intensitäten $\overline{F}_{h,k,P}$ und $\overline{F}_{v,k,P}$ und die Standardabweichungen ($\sigma_{h,k,P}$ und $\sigma_{v,k,P}$) berechnet.

Hintergrundsubtraktion

Die Segmentierung des Hintergrundes erfolgt für jedes Pixel eines aufgenommenen Bildes durch den Vergleich mit dem Hintergrundmodell. Dazu wird eine Schwellwertfunktion (22.2) auf jeden Kanal des Farbdifferenzbildes ΔI und der Kantendifferenzbilder ΔI_h und ΔI_v angewendet.

$$\Delta I = I - \overline{I}_{bck} = I - \frac{1}{n} \cdot \sum_{i=1}^{n} I_{bck,i} \tag{22.1}$$

$$C = \begin{cases} 0\% & , x < m \cdot \sigma \\ \frac{x - m \cdot \sigma}{(M-m) \cdot \sigma} \times 100\% & , m \cdot \sigma \leq x \leq M \cdot \sigma \\ 100\% & , x > M \cdot \sigma \end{cases} \tag{22.2}$$

Das Maximum aus so berechnetem Farb-Vordergrundbild C_{col} und Kanten-Vordergrundbild C_{edge} führt zu einer Grauwertmaske des Vordergrundes C_{max}.

3.2 Erweiterung und Adaption des grundlegenden Ansatzes

Für eine robuste Verifikation der korrekten Bauteilpositionierung ist entscheidend, dass der Vordergrund möglichst exakt vom Hintergrund segmentiert wird. Daher wird der als Basis verwendete Algorithmus durch nachfolgend näher beschriebene Ansätze erweitert, um eine robuste und möglichst exakte Segmentierung zu ermöglichen. Eine Übersicht über den erweiterten Algorithmus zur Hintergrundsubtraktion in seiner Gänze zeigt Abbildung 22.2.

Verbesserung der Farbsensitivität

Viele Verfahren, die sich in den zahlreichen veröffentlichten Evaluationen unabhängig von der Art ihrer Modellierung als sehr performant erwiesen haben (z.B. [10], [5]), nutzen im Rahmen ihres Farbmodells lediglich die Farbkanäle des RGB-Farbraumes. Die Verwendung eines metrischen Farbraumes, der beispielsweise auch von Barnich [11] verwendet wird, führt zu einer deutlichen Verbesserung der Hintergrundsubtraktion. Mit der Schwellwertfunktion (22.2) und $x = R_{col} \cdot \Delta I$ wird anschließend über alle drei Kanäle des Farbbildes die Zugehörigkeit C_{col} jedes Pixels zum Vordergrund klassifiziert. Gleiches erfolgt für das Kantenbild mit $x = R_{edge} \cdot \Delta I_{hv}$.

$$\Delta I_{hv} = |\overline{I}_{h,bck} - I_h| + |\overline{I}_{v,bck} - I_v| \tag{22.3}$$

$$R_{edge} = \frac{\Delta I_{hv}}{G^*_{edge,t}} \tag{22.4}$$

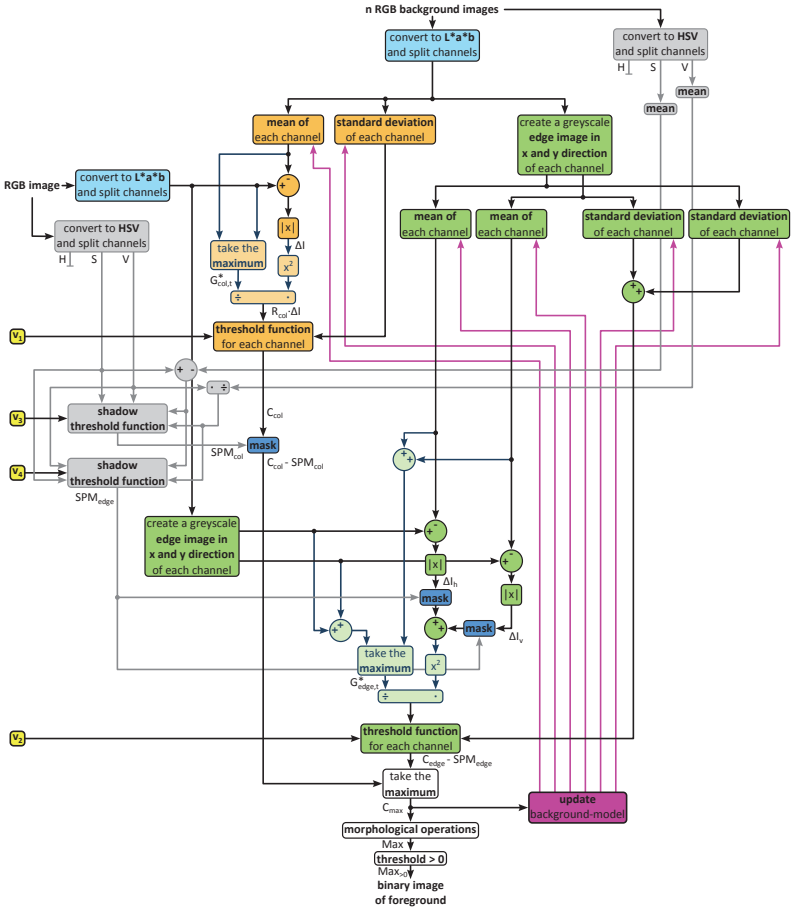


Abbildung 22.2: Struktur der vorgeschlagenen Hintergrundsubtraktion.

Die Segmentierung mittels der Schwellwertfunktion lässt sich gegenüber Jabri verbessern, indem die Parameter m und M in (22.2) (jeweils für Farb- und Kantenbild), bzw. die Parametersätze v_1 und v_2 in Abbildung 22.2, nicht statisch gesetzt, sondern adaptiv berechnet werden. Dazu bietet sich die Methode nach Rosin [12] an, um die der Algorithmus zur Hintergrundsubtraktion ergänzt wurde.

Verbesserung durch Schattendetektion

Ein weiterer Ansatzpunkt zur signifikanten Verbesserung der Hintergrundsubtraktion ist die Integration einer Schattenverbesserung. Dazu eignet sich das Verfahren nach Tattersall [13]. Dieses operiert ebenfalls in einem metrischen Farbraum und liefert ein binäres Bild mit als Schatten klassifizierten Pixeln, die anschließend im Farb- und im Kantenbild entfernt werden können. Der ursprüngliche Algorithmus nach Tattersall zeichnet sich insbesondere durch seine dynamische Parameterschätzung aus. Diese Parameter müssen für den statischen Anwendungsfall statisch gesetzt und empirisch auf die Umgebungsbedingungen abgestimmt werden.

3.3 Verifikation der korrekten Bauteilpositionierung

Die Verifikation der korrekten Positionierung eines Bauteils erfolgt durch Template Matching. Dazu wird der segmentierte Vordergrund eines Referenzbildes B_r mit dem des zu prüfenden Bildes B_w mittels multidirektionaler Kreuzkorrelation (vgl. [14]) verglichen.

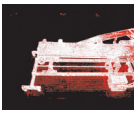
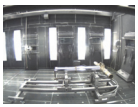
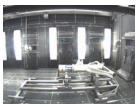
$$E_c = min\{B_r \star B_w, B_w \star B_r\} \tag{22.5}$$

Zum einen erhält man durch das Minimum beider Übereinstimmungswerte die grundsätzliche prozentuale Übereinstimmung von Referenz- und Prüfbild. Zum anderen liefert die Kreuzkorrelation auch die Bildkoordinaten des Zentrums der größten Übereinstimmung, die Rückschlüsse auf die globale Position der flexibel positionierbaren Bauteilaufnahme in der Roboterzelle erlauben.

4 Evaluation

Zur Evaluation werden in Tabelle 22.1 die Ergebnisse der Algorithmen zur Hintergrundsubtraktion nach Jabri mit denen des hier vorgestellten Ansatzes verglichen. Um die Robustheit gegenüber Reflexionen und Überbelichtungseffekten zu zeigen, wurden die Bauteile 2 und 3 überbelichtet aufgenommen. Erkennbar sind die signifikanten Auswirkungen der vorgeschlagenen Verbesserungen hinsichtlich fehlerhaft erkannter Pixel. Dies wird auch anhand der Tabelle 22.3 deutlich, die die

Tabelle 22.1: Hintergrundsubtraktion für drei exemplarische Bauteile.

	original	GT	Algorithmus nach Jabri	Erweiterter Algorithmus
Bauteil 1				
Bauteil 2				
Bauteil 3				

Fehlerhaft detektierte Pixel gegenüber GT sind rot markiert.

Genauigkeit der Algorithmen anhand dreier Verhältnisse $V1 - V3$ gegenüberstellt. Der Vergleich erfolgt jeweils gegenüber einem händisch freigestellten Referenzbild (GT).

$$V1 = \frac{\text{\#korrekt klassifizierte fg und bg Pixel}}{\text{\#Pixel des Gesamtbildes}} \tag{22.6}$$

$$V2 = \frac{\text{\#korrekt klassifizierte innerhalb GT}}{\text{\#Pixel innerhalb GT}} \tag{22.7}$$

$$V3 = \frac{\text{\#korrekt klassifizierte außerhalb GT}}{\text{\#Pixel außerhalb GT}} \tag{22.8}$$

Die Auswertungen zeigen, dass die Genauigkeit des vorgestellten Verfahrens unter günstigen sowie ungünstigen Bedingungen eine Genauigkeit korrekt erkannter Pixel von über 95% aufweist. Dies bildet die Grundlage für eine robuste Verifikation eines Bauteils und dessen Positionierung in der Roboterzelle durch Vergleich mittels Kreuzkorrelation.

Die Ergebnisse der Positionierungsverifikation sind in Tabelle 22.3 aufgeführt. Dazu wurden exemplarisch drei verschiedene Bauteile nach der vorgeschlagenen Methode verglichen. Es zeigt sich, dass gleiche Bauteile (bei gleicher Positionierung) zu einem Übereinstimmungswert von $> 95\%$ führen. Vergleiche mit anderen Bauteilen sehr ähnlicher

Kontur führen zu Werten $< 87\%$ (z. B. 1 vs. 2); Vergleiche mit Bauteilen anderer Kontur führen zu Werten $< 80\%$ (z. B. 1 vs. 3).

Tabelle 22.2: Qualitativer Vergleich des Algorithmus von Jabri mit dem erweiterten Algorithmus.

Bauteil	Verhältnis	Algorithmus nach Jabri	Erweiterter Algorithmus
1	V1	88,8%	96,2%
	V2	85,0%	97,2%
	V3	90,1%	95,8%
2	V1	91,6%	96,1%
	V2	82,1%	79,1%
	V3	93,3%	98,7%
3	V1	91,1%	96,2%
	V2	83,3%	81,5%
	V3	92,5%	98,5%

Tabelle 22.3: Template Matching für drei exemplarische Bauteile.

	Bauteil 1	Bauteil 2	Bauteil 3	Vordergrund
Bauteil 1	$96,0\%$ $\Delta_{ac} = 0$	$86,2\%$ $\Delta_{ac} = 0$	$74,3\%$ $\Delta_{ac} = 0$	
Bauteil 2	$86,2\%$ $\Delta_{ac} = 0$	$96,1\%$ $\Delta_{ac} = 0$	$74,1\%$ $\Delta_{ac} = 0$	
Bauteil 3	$74,3\%$ $\Delta_{ac} = 0$	$74,1\%$ $\Delta_{ac} = 0$	$95,5\%$ $\Delta_{ac} = 0$	

ac in [%], Abstand der ac-Werte Δ_{ac} in [pixel]

Die Auswirkungen eines korrekten Bauteils in der Roboterzelle, das sich in falscher Positionierung (translatorisch, rotatorisch) befindet, zeigt Tabelle 22.4. Eine translatorische Fehlpositionierung führt zum einen zu einer erkennbaren Abweichung des Übereinstimmungswertes sowie zu einer signifikanten Abweichung des Zentrums größter Übereinstimmung. Bereits kleine rotatorische Abweichung führen zu einer Reduktion des Übereinstimmungswertes. Anhand der vorgestellten Ergebnisse kann gezeigt werden, dass sowohl falsche Bauteile als auch korrekte Bauteile in falscher Positionierung robust erkannt werden können.

Tabelle 22.4: Template Matching für ein fehlerhaft positioniertes Bauteil.

Position	ac	Δ_{ac}	fg
unverändert	95,5%	0 pixel	
translatorischer Fehler (15cm)	78,0%	52 pixel	
rotatorischer Fehler (5°)	91,6%	0 pixel	

5 Zusammenfassung

Das vorgeschlagene Verfahren ermöglicht die Verifikation der korrekten Positionierung eines Bauteils für ein großes Bauteilspektrum von etwa 100-200 zu verifizierenden Bauteilen. Die in der Industrieumgebung durchgeführten Evaluationen zeigen die Robustheit des Verfahrens unter den äußerst suboptimalen Bedingungen sowie die Tauglichkeit zur Verwendung in Form eines Assistenzsystems. Ein solches Bildverarbeitungssystem lässt sich mit wenig Aufwand sowohl in eine Produktionsumgebung selbst als auch in Prozessführungssysteme integrieren. Die Informationen über ein neues zu verifizierendes Großbauteil lassen sich benutzerfreundlich durch einmalige Aufnahme von Referenzbildern erfassen.

Literatur

1. A. Sahay, *Leveraging information technology for optimal aircraft maintenance, repair and overhaul (MRO)*. Elsevier, 2012.
2. P. Ayeni, T. S. Baines, H. Lightfoot und P. Ball, „State-of-the-art of'lean'in the aviation maintenance, repair, and overhaul industry", *Proceedings of the Institution of Mechanical Engineers, Part B: journal of engineering manufacture*, S. 0954405411407122, 2011.
3. A. Sobral und A. Vacavant, „A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos", *Computer Vision and Image Understanding*, Vol. 122, S. 4–21, 2014.

4. S. Brutzer, B. Hoferlin und G. Heidemann, „Evaluation of background subtraction techniques for video surveillance“, in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, S. 1937–1944.
5. M. Hofmann, P. Tiefenbacher und G. Rigoll, „Background segmentation with feedback: The pixel-based adaptive segmenter“, in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. IEEE, 2012, S. 38–43.
6. S.-C. S. Cheung und C. Kamath, „Robust techniques for background subtraction in urban traffic video“, in *Proceedings of SPIE*, Vol. 5308, Nr. 1, 2004, S. 881–892.
7. M. Piccardi, „Background subtraction techniques: a review“, in *Systems, man and cybernetics, 2004 IEEE international conference on*, Vol. 4. IEEE, 2004, S. 3099–3104.
8. D. H. Parks und S. S. Fels, „Evaluation of background subtraction algorithms with post-processing“, in *Advanced Video and Signal Based Surveillance, 2008. AVSS'08. IEEE Fifth International Conference on*. IEEE, 2008, S. 192–199.
9. S. Herrero und J. Bescós, „Background subtraction techniques: systematic evaluation and comparative analysis“, in *Advanced Concepts for Intelligent Vision Systems*. Springer, 2009, S. 33–42.
10. S. Jabri, Z. Duric, H. Wechsler und A. Rosenfeld, „Detection and location of people in video images using adaptive fusion of color and edge information“, in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, Vol. 4. IEEE, 2000, S. 627–630.
11. O. Barnich und M. Van Droogenbroeck, „Vibe: a powerful random technique to estimate the background in video sequences“, in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. IEEE, 2009, S. 945–948.
12. P. L. Rosin, „Unimodal thresholding“, *Pattern recognition*, Vol. 34, Nr. 11, S. 2083–2096, 2001.
13. S. Tattersall und K. Dawson-Howe, „Adaptive shadow identification through automatic parameter estimation in video sequences“, *Proceedings of Irish Machine Vision and Image Processing*, S. 57–64, 2003.
14. J. C. Russ, *The image processing handbook*. CRC press, 2010.

Kamerabasiertes Referenzsystem für Fahrerassistenzsysteme

Tim Kubertschak, Marina Wittenzellner und Mirko Maehlisch

AUDI AG, Konzeptentwicklung Pilotiertes Fahren,
D-85045 Ingolstadt

Zusammenfassung In diesem Beitrag wird ein neuartiges kamerabasiertes System zur Referenzierung von Fahrerassistenzsystemen vorgestellt. Das System basiert auf der Erkennung und Verarbeitung spezieller Markierungen an allen beteiligten Fahrzeugen, um eine eindeutige Identifikation zu ermöglichen. Im Beitrag wird die Struktur der zu referenzierenden Szene beschrieben, alle notwendigen Transformationen zur Positionsbestimmung hergeleitet und eine Analyse der erreichbaren Genauigkeit durchgeführt.

1 Einleitung

Die aktuellen Entwicklungen in der Automobilindustrie werden derzeit maßgeblich durch Systeme geprägt, die den Fahrer in unterschiedlichen Situationen unterstützen. Von besonderem Interesse sind kritische Situationen in denen der Fahrer nicht mehr reagieren kann oder für die Situation ungünstig reagiert. Nach dem Willen der Hersteller mündet dieses Bestreben in einem komplett autonomen Betrieb des Fahrzeugs, der keine Eingriffe des Fahrers mehr benötigt. Um dieses Ziel sowohl technisch als auch rechtlich zu erreichen, müssen die Systeme umfassend abgesichert werden. Der erwartete Aufwand ist dabei um ein Vielfaches höher als es bei heutigen Systemen der Fall ist.

Für die Absicherung von Fahrerassistenzsystemen müssen unterschiedliche Objekte im Umfeld des Fahrzeugs durch ein Referenzsystem erkannt und mit den Ergebnissen der Assistenzfunktionen verglichen werden. Je nach System muss sowohl auf statische Hindernisse (Bordsteine, Hauswände, Fahrbahnmarkierungen) als auch auf dyna-

mische Hindernisse (Fahrzeuge, Fußgänger, Fahrräder) reagiert werden. Die Reaktion muss je nach abzusicherndem System und der Kritikalität der Situation sehr schnell erfolgen. Das setzt eine hohe Genauigkeit der Umfeldrepräsentation und eine hohe Genauigkeit des Referenzsystems voraus.

Zur Erzeugung von Referenzdaten werden derzeit hauptsächlich zwei unterschiedliche Verfahren verwendet. Sie bestehen aus der Kombination von hoch genauer Positionsbestimmung und Vermessung des Umfeldes. Die Position wird im Allgemeinen durch spezielle GPS-basierte Systeme (z.B. Differential GPS, RTK-GPS) durchgeführt [1]. Die Vermessung der Umgebung erfolgt durch einen drei-dimensional messenden Laserscanner (3D-LiDAR) [2] oder durch manuelles bzw. semimanuelles Labeln von Messdaten [3]. Im ersten Verfahren wird eine sehr große Datenmenge erzeugt, deren Information im Anschluss auf intelligente Art und Weise komprimiert wird oder aus welcher durch komplexe Algorithmen die interessanten Objekte extrahiert werden. Das Ergebnis müssen im Anschluss stichprobenartig gesichtet werden, um ein Fehlverhalten der Algorithmen auszuschließen. Beim zweiten Verfahren werden relevante Objekte in den einzelnen Datenframes exakt markiert um Referenzdaten zu generieren. Die Markierung erfolgt entweder manuell in jedem Datenframe oder im jeweiligen Startframe mit algorithmischer Verfolgung über alle weiteren relevanten Frames.

Diese Methoden können jedoch nicht direkt auf die Absicherung zukünftiger Fahrerassistenzsysteme übertragen werden. Für den erforderlichen Umfang der Absicherung ist der erzeugte Datensatz zu groß oder fordert zu großen Aufwand bei der Nachbearbeitung. Zudem sollten die einzelnen Module zukünftiger Assistenzsysteme einer stärkeren entwicklungsbegleitenden Referenzierung unterzogen werden, um die abschließende Absicherung zu vereinfachen. Dafür muss das Referenzsystem einfach einzusetzen sein, damit die Entwickler der Komponenten die Möglichkeit der Referenzierung häufig nutzen. Der zeitliche Aufwand zum Einsatz der beschrieben Systeme, insbesondere zur Kalibrierung [4, 5] lassen einen entwicklungsbegleitenden Einsatz aber kaum zu.

Der Einsatz eines 3D-LiDAR zur Referenzierung ist außerdem bezüglich der funktionalen Sicherheit von Fahrerassistenzfunktionen problematisch. Der Standard ISO 26262 [6] zur funktionalen Sicherheit von Fahrzeugen fordert für besonders sicherheitskritische Funktionen,

dass bestimmte Ereignisse durch redundante Sensorik ausgelöst werden müssen. Die Redundanz bezieht sich dabei auf die Messprinzipien der Sensoren. So gilt das Eintreten eines Ereignisses nicht als funktional abgesichert, wenn es von zwei Ultraschallsensoren hervorgerufen wurde. Gleichzeitig werden LiDAR-Systeme für den Einsatz in zukünftigen Assistenzfunktionen evaluiert [7]. Vor diesem Hintergrund lässt sich ein alleiniger Einsatz eines 3D-LiDAR zur Referenzierung kaum begründen.

In diesem Beitrag wird zur Lösung der angesprochenen Probleme ein neuartiges System vorgestellt, welches Kameradaten zu Referenzierung von Fahrerassistenzsystemen verwendet. Es basiert auf der Detektion und Verarbeitung spezieller Markierungen über die alle beteiligten Verkehrsteilnehmer jederzeit eindeutig zu identifizieren sind. Die einzelnen Schritte beginnend bei der Datenaufnahme bis zur Lokalisierung aller bekannten relevanten Teilnehmer wird in Abschnitt 2 beschrieben. Die theoretisch erreichbaren Genauigkeiten werden in Abschnitt 3 hergeleitet und praktisch validiert. Eine beispielhaftes Anwendung des Referenzsystem wird in Abschnitt 4 gezeigt, bevor die einzelnen Ergebnisse des Beitrag in Abschnitt 5 zusammengefasst werden.

2 Kamerabasiertes Referenzsystem

Zur Referenzierung wird ein bildbasiertes bzw. kamerabasiertes System vorgeschlagen. Die Bildaufnahme erfolgt aus der Vogelperspektive, um alle für Fahrerassistenzsysteme relevanten geometrischen Eigenschaften der beteiligten Objekte abzubilden. Von besonderem Interesse ist die laterale Ausdehnung der Objekte, die für die Überprüfung der Genauigkeiten einzelner Komponenten von Assistenzsystemen wichtig ist. Die Identifikation erfolgt über spezielle Markierungen, die auf der Oberfläche der Objekte angebracht sind. Für die Ableitung der Referenzdaten aus diesen Markierungen wird zusätzliches Modellwissen bezüglich der Szene verwendet.

In den drei folgenden Abschnitten werden alle notwendigen Schritte beschrieben, um aus dem Kamerabild Referenzdaten zu gewinnen. Dazu zählt der Aufbau der Szene zur Positionsbestimmung aller Teilnehmer, die Beschreibung der Marker zur Identifikation und die Kette der Verarbeitungsschritte.

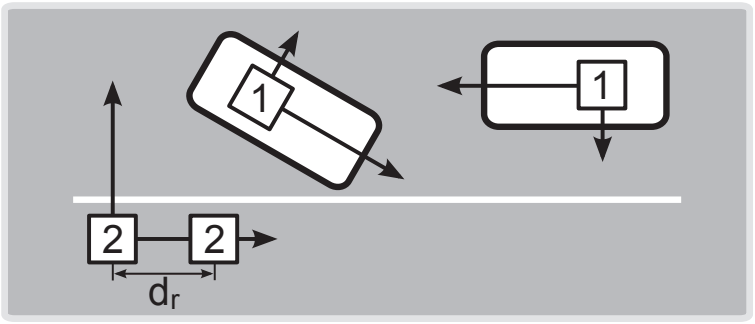


Abbildung 23.1: Beschreibung einer Szene des Referenzsystems. Zu sehen sind zwei Verkehrsteilnehmer mit ihren lokalen Koordinatensysteme, sowie das Referenzkoordinatensystem.

2.1 Szenendefinition

Für einen problemlosen Ablauf der Referenzierung einzelner Szenen müssen bestimmte Rahmenbedingungen eingehalten werden. Wie in Abbildung 23.1 zu sehen ist, besteht eine vollständige Szene aus zwei Typen von Markern. Durch die Markierungen vom Typ 1 werden alle an der Szene beteiligten Teilnehmer festgelegt. Die beiden Marker vom Typ 2 werden genutzt, um innerhalb der Szene ein Referenzkoordinatensystem zu definieren. Die Abszisse verläuft durch die Mittelpunkte der beiden Marker, die Ordinate senkrecht zu ihr und schneidet sie im Mittelpunkt des ersten Markers. Über den Abstand d_r zwischen den beiden Markierungen wird der Skalierungsfaktor und damit auch die Auflösung der Kamera bestimmt. Dadurch wird jeder Teilnehmer unabhängig von der Position der Kamera eindeutig in der Szene lokalisiert.

2.2 Markerbeschreibung

Zur Identifikation der Teilnehmer während der Referenzierung und zur Definition des Referenzkoordinatensystems werden spezielle Markierungen verwendet. Die Markierungen werden bevorzugt auf den Dächern der Teilnehmer angebracht, damit sie von erhöhter Kameraposition gut zu erkennen sind. Ein Beispiel der verwendeten Marker

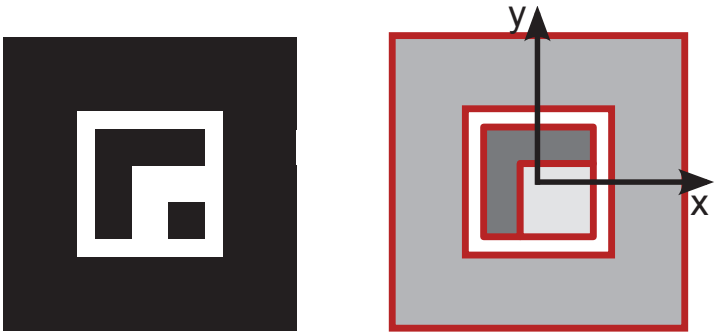


Abbildung 23.2: Beispielhafter Marker sowie dessen Struktur bestehend aus drei Bereichen. Der Markerursprung befindet sich in Mitte des Markers.

und deren Struktur ist in Abbildung 23.2 gezeigt. Wie zu erkennen ist, weißen sie eine starke Ähnlichkeit zu DataMatrix-Codes [8] auf.

Die Marker bestehen aus drei verschiedenen Bereichen, die im rechten Teil von Abbildung 23.2 hervorgehoben sind. Der große äußere Bereich ist für eine stabile Erkennung der Markierungen notwendig. Er wird durch das verwendete Verfahren vorgeschrieben (siehe Abschnitt 2.3) und nimmt etwa die Hälfte des kompletten Markers ein. Der dunkle innere Bereich dient zur Detektion der Orientierung des Markers im Kamerabild. Durch seine Form, kann jederzeit eine eindeutige Aussage über die Ausrichtung des Markers getroffen werden und damit über die Ausrichtung des Objektes. Er nimmt etwa die Hälfte des inneren Bereich ein. Die anderen Hälfte wird durch den Datenbereich belegt. Er dient der Unterscheidung verschiedener, im Bild befindlicher Objekte, indem jedem Objekt ein bestimmtes Codewort zugeordnet wird. Die Codeworte bestehen wie bei anderen Markern aus einzelnen Bits (d.h. schwarzen und weißen Quadraten). Im Unterschied zu diesen Markern ist allerdings keinerlei Redundanz in der Kodierung notwendig, da während der Referenzierung weder mit Verdeckungen noch mit Beschädigungen der Marker zu rechnen ist. Der Verzicht erlaubt die Verwendung eines minimalen Codes, wodurch der Code eine große Fläche belegen kann und damit eine gute Detektion ermöglicht wird.

2.3 Referenzdatenerstellung

Die Verarbeitung der Kameradaten folgt in den ersten Schritten klassischen Bildverarbeitungsansätzen. Nach der Bildaufnahme erfolgt eine Rektifizierung des Bildes, um jegliche Linsenverzerrungen zu beseitigen. Anschließend werden die Marker mit dem von Kato und Billinghurst [9] vorgeschlagenen Verfahren detektiert. Dabei wird der im Abschnitt 2.2 beschriebene spezielle Aufbau der Marker ausgenutzt. Die Detektion der Position erfolgt mit Hilfe des breiten äußeren Bereichs. Dieser stellt ein in natürlichen Szenen selten vorkommendes Muster dar, so dass die systematische Suche nach diesen Mustern ein ausreichendes Unterscheidungsmerkmal zur Umgebung darstellt. Der innere Bereich wird nach der Detektion des Markers gemäß eines festen Raster abgetastet, um eine entzerrte Version des inneren Bereichs in niedriger Auflöung zu erhalten. Diese wird mit einem angelernten Muster korreliert, um die Marker zu klassifizieren. Als Ergebnis dieses Schrittes liegt für jedes bekannte, im Bild enthaltene Objekt die Position und Ausrichtung des zugehörigen Markers in Bildkoordinaten vor. Die Extraktion der Position erfolgt subpixelgenau.

Um aus diesen bildbezogenen Informationen Referenzdaten zu gewinnen, wird eine Folge von Transformationen

$$\mathbf{T}_{Ref}^{Obj} = \underbrace{\mathbf{R}\left(\varphi_3\right) \cdot \mathbf{T}\left(s\Delta\mathbf{x}_3\right)}_{\mathbf{T}_3} \cdot \underbrace{\mathbf{T}\left(s\Delta\mathbf{x}_2\right) \cdot \mathbf{R}\left(\varphi_2\right)}_{\mathbf{T}_2} \cdot \underbrace{\mathbf{R}\left(\varphi_1\right) \cdot \mathbf{T}\left(\Delta\mathbf{x}_1\right)}_{\mathbf{T}_1} \quad (23.1)$$

bestehend aus Rotations- und Translationsmatrizen $\mathbf{R}$ und $\mathbf{T}$ mit entsprechenden Parametern auf die Objekte angewendet. Das Ziel dieser Transformationen ist die Abbildung der Objekte im lokalen Referenzkoordinatensystem. Die Transformation $\mathbf{T}_1$ überführt im ersten Schritt das objektlokale Koordinatensystem in Markerkoordinaten. Die Parameter der Matrizen sind die Position $\Delta\mathbf{x}_1$ und die Ausrichtung φ_1 des Markers bezüglich des Ursprungs der Objektkoordinaten. Durch $\mathbf{T}_2$ werden die Markerkoordinaten in normierte Bildkoordinaten transformiert. Dafür wird die Position $\Delta\mathbf{x}_2$ und Ausrichtung φ_2 des Markers im Bild verwendet. Die Position wird mit der Auflösung s normiert, da sie ausschließlich in Pixelkoordinaten vorliegt. Die Auflösung wird aus dem Abstand der beiden Referenzmarker in Pixelkoordinaten und dem Abstand d_r gewonnen. Mit $\mathbf{T}_3$ werden die normierten Bildkoordinaten schließlich in Referenzkoordinaten transformiert. Als Ausrichtung φ_3

und Position $\Delta \mathbf{x}_3$ werden jene der Referenzmarker verwendet. Da auch sie in Pixelkoordinaten vorliegen, müssen sie mit s normiert werden.

Zur Ableitung der Referenzdaten müssen schließlich noch die zwei folgenden Schritte durchgeführt werden. Einerseits sind Referenzdaten im lokalen Referenzkoordinaten nicht sehr nützlich, da dieses Koordinatensystem den einzelnen Fahrzeugen im Allgemeinen nicht bekannt ist. Andererseits bestehen die relevanten Objekt bisher nur aus einzelnen Positionen. Daher werden alle Objekte in das objektlokale Koordinatensystem eine bestimmten Objektes, dem sogenannten Ego-Objekt transformiert. Mathematisch entspricht das der Transformation $\mathbf{T}_{Ref}^{Ego^{-1}} \cdot \mathbf{T}_{Ref}^{Obj}$. Zudem werden die geometrischen Abmaße aller relevanten Objekte als Vorwissen dem Referenzsystem bereitgestellt. Der generelle Einsatz des Systems ist dadurch zwar eingeschränkt, für Fahrerassistenzsysteme aber unproblematisch da entsprechende Fahrzeugdaten im Allgemeinen vorhanden sind.

3 Genauigkeit des Referenzsystems

In den folgenden Abschnitten wird die theoretisch erreichbare Genauigkeit basierend auf typischen Fehlern hergeleitet. Diese wird im darauf folgenden Abschnitt praktisch validiert.

3.1 Theoretische Ableitung

Die theoretische Ableitung der Genauigkeit erfolgt nach dem Gaußschen Fehlerfortpflanzungsgesetz [10]. Jeder Messwert wird als normal verteilte statistische Größe aufgefasst, die mit einer bestimmten Varianz σ^2 um den realen Wert rauscht. Die Messunsicherheiten pflanzen sich durch alle Verarbeitungsschritte fort. Die Varianz des Ergebnisses ergibt sich schließlich als gewichtete Summe der Varianzen der Messwerte. Als Gewicht wird der Einfluss des Messwertes am Ergebnis verwendet, der dem Quadrat der partiellen Ableitung der Verarbeitungsvorschrift nach dem Messwert entspricht.

Die Transformationsvorschrift aus Gleichung (23.1) lässt sich durch die Funktionen

$$\begin{pmatrix} f_x \\ f_y \end{pmatrix} = \mathbf{T}_{Ref}^{Obj} \cdot \begin{pmatrix} x \\ y \end{pmatrix} \tag{23.2}$$

darstellen. f_x und f_y entsprechen dabei den Gleichungen zur Berechnung der transformierten x- bzw. y-Koordinate. Mit Gleichung (23.2) folgt für die Varianz der transformierten Koordinate

$$\begin{pmatrix} \sigma^2_{\mathbf{T}(x)} \\ \sigma^2_{\mathbf{T}(y)} \end{pmatrix} = \mathbf{J} \cdot \left(\sigma^2_{\varphi_1}\ \sigma^2_{\Delta\mathbf{x}_1}\ \sigma^2_{\varphi_2}\ \sigma^2_{\Delta\mathbf{x}_2}\ \sigma^2_{\varphi_3}\ \sigma^2_{\Delta\mathbf{x}_3}\ \sigma^2_s\right)^T \cdot \mathbf{J}^T, \tag{23.3}$$

wobei $\mathbf{J}$ die Jacobimatrix der Funktionen f_x und f_y und der Unbekannten φ_1, $\Delta\mathbf{x}_1$, φ_2, $\Delta\mathbf{x}_2$, φ_3, $\Delta\mathbf{x}_3$ und s ist und die σ_i^2 die jeweiligen Messvarianzen sind.

3.2 Validierung

Die im letzten Abschnitt hergeleiteten Genauigkeiten werden in diesem Abschnitt validiert. Dafür wurde eine Testszene vorbereitet, in der die Positionen aller relevanten Teilnehmer zueinander exakt vermessen wurden. Für die Szene wurden mit dem vorgeschlagenen System Referenzdaten erzeugt und die Abweichungen einiger charakteristischer Punkte von der Realität bestimmt. Die Wahl der Punkte basiert auf zwei Kriterien: Einerseits müssen sie sich an der Fahrzeugkontur abheben, andererseits sollte der Einfluss durch Unsicherheiten unterschiedlich stark sein. Die Testszene mit den gewählten Testpunkten ist in Abbildung 23.3 dargestellt. Die Genauigkeit der einzelnen Punkte wird insbesondere durch die Winkelunsicherheit unterschiedlich stark beeinflusst, wie die Validierung zeigen wird.

Die Szene wurde mit der Kamera UI-1240SE-M-GL von IDS aus einer Höhe von 16 m aufgenommen. Bei dieser Konfiguration ist eine Auflösung von etwa 20 mm erreichbar. Das heißt, die Positionen der Marker können bei feststehender Kamera durch Variationen der Itensitäten innerhalb der Auflösung, also um maximal einen Pixel schwanken. Zur Vereinfachung wird angenommen, dass die Unsicherheiten an allen Stellen des Bildes identisch sind. Somit werden die Abweichungen $\sigma_{\Delta\mathbf{x}_2}$ und $\sigma_{\Delta\mathbf{x}_3}$ auf 10 mm gesetzt. Für σ_{d_r} und $\sigma_{\Delta\mathbf{x}_1}$ wird 2,5 mm angenommen, die Winkelunsicherheiten σ_{φ_i} werden auf 0.1° gesetzt. Die Ergebnis der Validierung und die theoretisch erreichbare Genauigkeit sind im Box-Whisker-Plot in Abbildung 23.4 angegeben.

Wie der Vergleich mit dem rechten Plot in der Abbildung zeigt, liegen die praktisch ermittelten Abweichungen sehr nah an der theoretisch

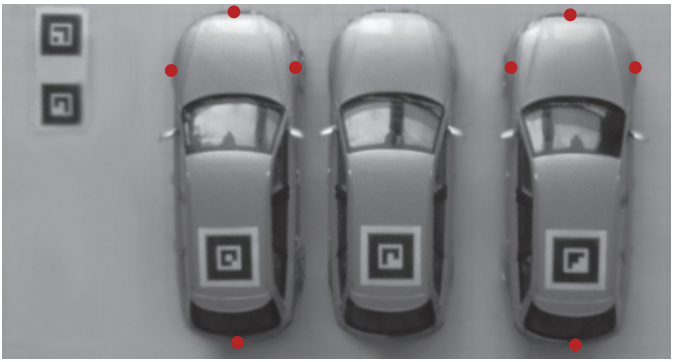


Abbildung 23.3: Szene zur Validierung des Referenzsystems. Rot markiert sind charkteristische Punkte, anhand deren die Genauigkeit verifiziert wurde. Das Ego-Fahrzeug ist jenes in der Mitte.

möglichen Genauigkeit bei etwa 10 mm. Dieser Wert ist sehr gut und erreicht an die Genauigkeiten anderer Referenzsysteme bzw. verbessert sie. Bei einigen Testpunkten konnte außerdem eine deutlich bessere Genauigkeit als die theoretisch ermittelte erreicht werden. Zu erklären ist dieser Effekt durch zwei Dinge: Einerseits werden die Markerpositionen wie in Abschnitt 2.3 beschrieben subpixelgenau extrahiert. Andererseits scheinen die Testpunkte nicht in gleicher Weise durch die unterschiedlichen Unsicherheiten beeinflusst zu werden.

4 Anwendung

Als Anwendung des vorgestellten Systems wird in diesem Abschnitt die Erzeugung von Referenzdaten für ein einfaches Parkszenario präsentiert. Das Szenario dient als Test für einen Parkassistenten. An dem Szenario sind drei Fahrzeuge beteiligt, wie auf der linken Seite in Abbildung 23.5 zu erkennen ist. Die Fahrzeuge mit den Nummern 1 und 3 sind bereits geparkt. Das Fahrzeug mit der Nummer 2 soll zwischen den beiden anderen Fahrzeugen eingeparkt werden, wofür eine exakte Kenntnis des Bereichs notwendig ist.

Wie in der Abbildung zu erkennen ist, werden alle relevanten Teilnehmer der Szene exakt wiedergegeben. Die Wiedergabe erfolgt in die-

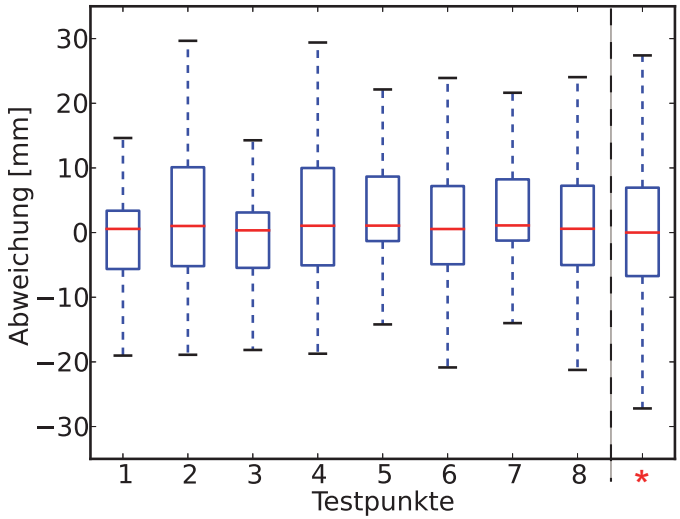


Abbildung 23.4: Box-Wister-Plot der Fehlerverteilungen einiger ausgewählter Punkte auf der Fahrzeugkontur der in Abbildung 23.3 gezeigten Szene. Der mit dem roten Stern markierte Plot zeigt die theoretisch erreichbare Genauigkeit.

Abbildung 23.5: Testszenario für das Referenzsystem (links) und die daraus abgeleiteten Referenzdaten (rechts). Die kritischen Verkehrsteilnehmer sind durch ihre Konturen in Form einer topologischen Abbildung der Umgebung [11] dargestellt. Das Ego-Fahrzeug (2) ist ausschließlich zur besseren Verständlichkeit dargestellt. Es gehört nicht zu den Referenzdaten.

sem Beispiel durch eine topologischen Beschreibung [11] des Fahrzeugumfeldes. Die Genauigkeit einer Komponente des angesprochenen Parkassistenten lässt sich schließlich durch einen Vergleich der topologischen Beschreibungen bewerten.

5 Zusammenfassung und Ausblick

In diesem Beitrag wurde ein neuartiges System zur Referenzierung von Fahrerassistenzsystemen basierend auf Kamerabildern und speziellen Markierungen an allen Teilnehmern vorgestellt. Zur Ableitung von Referenzdaten wurden alle notwendigen Komponenten beschrieben. Diese Komponenten sind der Szenenaufbau inklusive der Definition eines bildunabhängigen Koordinatensystems, der Aufbau der Markierungen und die notwendige Bildverarbeitung. Zur Bewertung des vorgeschlagenen Referenzsystems wurden verschiedene Analysen zur Genauigkeit durchgeführt. Neben einer theoretischen Ableitung der Genauigkeit erfolgt eine Validierung anhand realer Daten, wobei die Positionen einiger charakteristischer Punkte über einen längeren Zeitraum verfolgt wurde. Die Validierung ergab ähnliche Positionsfehler wie die theoretische Bestimmung. Die mittlere Abweichung beträgt für die verwendete Konfiguration aus Kameraauflösung und -position lediglich 10 mm.

Das Referenzsystem, so wie in diesem Beitrag beschrieben, ist in seiner Einsatzmöglichkeit beschränkt. Es können derzeit nur relativ kleine Szenen mit einem begrenzten Bewegungsradius der Fahrzeuge referenziert wird. Die Ursache dafür ist die statische Position der Kamera. Eine sinnvolle Weiterentwicklung des Systems ist daher die Erweiterung auf größere Szenen. Eine Möglichkeit die dafür untersucht werden soll, ist die Befestigung der Kamera an Quadrocoptern oder Drohnen. Als Herausforderung muss die Stabilisierung des Bildes und die Definition eines bildunabhängigen Koordinatensystems gelöst werden.

Literatur

1. R. Grewe, A. Hohm, S. Lüke und M. Komar, „Genauigkeitsanalyse eines gridbasierten Verfahrens zur Umfeldbeschreibung“ , in *Proceedings of 5th Tagung Fahrerassistenz*, 2012.

2. J. Klandt, M. Radimirsch, A. Kirchner und H. Philipps, „Referenzierung von Fahrerassistenzsystemen: Mess- und Auswertesysteme für die Umfelderfassung“, in *Proceedings of 28. VDI/VW-Gemeinschaftstagung Fahrerassistenz und Integrierte Sicherheit*, Ser. VDI-Berichte, Nr. 2166. VDI Verlag GmbH, 2012, S. 215 – 225.
3. M. Benmimoun, F. Fahrenkrog, A. Pütz, A. Zlocki und L. Eckstein, „Wirkungsanalyse von ACC und FCW auf Grundlage von CAN-Daten im Rahmen eines Feldversuchs“, in *Proceedings of 8th Workshop Fahrerassistenzsysteme*, K. Dietmayer, Hrsg., 2012, S. 27 – 37.
4. J. Levinson und S. Thrun, „Unsupervised calibration for multi-beam lasers“, in *Proceedings of 12th International Symposium on Experimental Robotics*, Ser. Springer Tracts in Advanced Robotics, O. Khatib, V. Kumar und G. Sukhatme, Hrsg., Nr. 79. Springer Berlin Heidelberg, 2014, S. 179 – 193.
5. S. Hong, M. H. Lee, H.-H. Chun, S.-H. Kwon und J. L. Speyer, „Experimental study on the estimation of lever arm in GPS/INS“, *IEEE Transactions on Vehicular Technology*, Vol. 55, Nr. 2, S. 431 – 448, March 2006.
6. ISO/TC 22/SC 3, *Roadvehicles – Functional Safety*, International Organization for Standardization, ISO 26262: 1 – 10, 2012.
7. S. Gidel, P. Checchin, C. Blanc, T. Chateau und L. Trassoudaine, „Pedestrian detection and tracking in urban environment using a multilayer laserscanner“, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 11, Nr. 3, S. 579 – 588, September 2010.
8. ISO/IEC JTC 1/SC 31, *Information technology – Automatic identification and data capture techniques – Data Matrix bar code symbology specification*, International Organization for Standardization, ISO/IEC 16022, 2006.
9. H. Kato und M. Billinghurst, „Marker tracking and HMD calibration for a video-based augmented reality conferencing system“, in *Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality*, 1999, S. 85 – 94.
10. P. R. Bevington und D. K. Robinson, *Data Reduction and Error Analysis*, 3. Aufl. McGraw-Hill Higher Eduction, 2003.
11. T. Kubertschak, M. Maehlisch und H.-J. Wuensche, „Fences - a unified architecture for mapping static environment for driver assistance systems“, in *Proceedings of 9th Workshop Fahrerassistenzsysteme*, B. Färber, Hrsg., 2014, S. 105 – 114.

Richtungsabhängige Personendetektion und -verfolgung

Johannes Pallauf und Fernando Puente León

Karlsruher Institut für Technologie (KIT),
Institut für Industrielle Informationstechnik (IIIT),
Hertzstr. 16, 76187 Karlsruhe

Zusammenfassung Dieser Beitrag untersucht die Nutzung von Orientierungsinformationen aus Einzelbildern zur Personenverfolgung. In dem vorgestellten Verfahren werden Farb- und Tiefeninformationen einer Kamera genutzt, um eine schnelle Detektion von Personen zu ermöglichen. Die dafür herangezogenen Merkmale werden anschließend weiterverwendet, um eine Klassifikation in vier Hauptrichtungen durchzuführen, die für eine Orientierungsbestimmung genutzt werden. Die erhaltene Orientierungs- und Positionsinformation wird zuletzt in einem auf dem Unscented-Kalman-Filter basierenden Verfolgungsalgorithmus fusioniert und evaluiert.

1 Einleitung

Die Verfolgung einer unbekannten Anzahl an Objekten in einem Zustandsraum hat zum Ziel, mit Sensoren die Bewegungspfade von Objekten, wie beispielsweise Fußgängern, zu erfassen. Hierzu werden, je nach Einsatzgebiet, verschiedene Sensoren wie Laserscanner, Kameras oder Radare eingesetzt. Neben der Extraktion einzelner Detektionen aus den Messdaten stellt die Fusion der Objekthypothesen verschiedener Sensoren eine weitere Herausforderung dar. Während oft die Anzahl der zu schätzenden Zustände, wie beispielsweise im Ein-Objekt-Fall, bekannt ist, muss im Multi-Objekt-Fall zusätzlich noch aus den Messdaten über die Anzahl der vorhandenen Objekte und somit die Menge der Zustände entschieden werden. Für effiziente Methoden sind in einer Vielzahl von Szenarien, wie Robotik, Fahrassistenzsysteme oder Überwachungseinrichtungen, Anwendungen denkbar. Für viele dieser

Einsatzbereiche spielen Echtzeitfähigkeit und somit ein überschaubarer Rechenaufwand eine wichtige Rolle.

Die Fusion von Daten zur Schätzung von Zuständen ist aufgrund der großen Zahl an Anwendungsgebieten ein vielbeachtetes Thema. Doch trotz intensiver Forschung sind heutige Systeme der Informationsfusion noch weit von der Leistung des menschlichen Gehirns entfernt.

Im Bereich der Bildverarbeitung stellt die Detektion von Personen gerade in komplexen Situationen ein weiterhin interessantes Forschungsgebiet dar. Während die meisten aktuellen Lösungen auf der Klassifikation von „*Histogram of Oriented Gradients* (HOG)"-Merkmalen [1–3] beruhen und nur zur Detektion dienen, gibt es darüber hinausgehende Ansätze, die zusätzlich versuchen, Informationen über die Orientierung zu gewinnen [4, 5], die hier als Grundlage dienen sollen.

Gerade die gleichzeitige Schätzung von Anzahl und Zustand der Objekte stellt eine erhöhte Anforderung an die Sensordatenverarbeitung dar. Es ist nötig, möglichst viel Information aus den Sensordaten zu gewinnen, um Zustände besser schätzen zu können. Gerade bei Multi-Objekt-Problemen ist in Szenarien, in denen sich Objekte auf engem Raum bewegen, eine Zuordnung von Messung zu Objekten nicht zweifelsfrei möglich, da oft nur die Position einer Messung bekannt ist. Des Weiteren sind gerade bei der Verfolgung von Personen übliche Bewegungsmodelle einer konstanten Geschwindigkeit oder Beschleunigung lediglich eine grobe Annäherung des tatsächlichen Bewegungsverhaltens von Fußgängern. Richtungsänderungen oder das korrekte Initialisieren von Objektorientierungen im Sensorsichtfeld müssen also aus der Schätzung der Trajektorie über der Zeit erkannt werden, die jedoch oft auf einem nur näherungsweise zutreffenden Modell beruht und zusätzlich weiteren Messunsicherheiten unterliegt.

Der Lösungsansatz der vorliegenden Arbeit beruht auf der Verwendung eines auf der Unscented-Kalman-Filterung [6] basierenden Filters. Dabei sollen als erweiterte Messungen Detektionen eines RGB-D-Sensors zur Farb- und Tiefenmessung dienen, die neben der zum Sensor relativen Personenposition aus den Tiefendaten auch eine Aussage über die relative Orientierung aus Einzelbildern liefern sollen. Für die Verarbeitung von Orientierungsinformationen wird ausgehend von Ergebnissen aus [4] ein auf HOG-Merkmalen basierter Personendetektionsalgorithmus für die Richtungserkennung erweitert. Dabei werden für durch Tiefendaten nach [3] ausgewählte Kandidatenbereiche die zu-

gehörigen Farbbilder mittels mehrerer *Support Vector Machines* (SVM) klassifiziert. Die Ergebnisse der vier Klassifikatoren der vier Hauptrichtungen können daraufhin als kontinuierliche Schätzung für die Zustandsschätzung genutzt werden.

Im Folgenden werden aufbauend auf dem Verfahren zur Personendetektion (Abschnitt 2) die vorgenommenen Erweiterungen der Orientierungsbestimmung (Abschnitt 3) erläutert. Darauf folgend wird das Modell für den Einbezug der zusätzlichen Messung in das Verfolgungsfilter vorgestellt (Abschnitt 4). Der Beitrag endet mit den Ergebnissen (Abschnitt 5) und einer Zusammenfassung (Abschnitt 6).

2 Personendetektion

Um Personen nicht nur in einem Videobild, sondern auch in einem Weltkoordinatensystem verfolgen zu können, sind Sensoren nötig, die über die reine Farbinformation hinaus noch Information über die Tiefe der beobachteten Szene liefern und somit Objekte im Raum lokalisieren können. Durch ihre günstige Verfügbarkeit bietet sich die Verwendung von RGB-D-Sensoren wie der Microsoft Kinect an. Die verfügbaren Tiefendaten bieten sich neben der Lokalisierung auch dazu an, eine Personendetektion zu verbessern und den hierfür nötigen Rechenaufwand gegenüber einem klassischen, rechenintensiven *Sliding-Window*-Ansatz aller möglicher Detektionsfenster [1] zu reduzieren [2, 3]. Der hier vorgestellte Ansatz basiert auf der Verwendung von HOG-Merkmalen, lässt sich jedoch auch auf andere Merkmale wie in [5] anwenden. Eine Übersicht über das vorgestellte Verfahren ist in Abbildung 24.1 zu sehen.

Zu Beginn des Detektionsschritts werden die Rohdaten in Form einer 3D-Punktwolke durch den RGB-D-Sensor erfasst. Um eine Reduktion des Suchraums zu erreichen, werden nun nach dem Vorbild von [3] zuerst alle Punkte, die zum Boden gehören und somit Verbindungen zwischen verbleibenden Punkten bilden, entfernt, was leicht unter der Annahme einer konstanten und bekannten Ausrichtung des Sensors zum Boden durchgeführt werden kann. Nun werden die verbleibenden Punktwolken aufgrund der Abstände zwischen den Punkten zu Gruppen zusammengefasst, die nun im Innenbereich beispielsweise zu detektierende Personen, Möbel oder Wände darstellen. Durch eine grobe

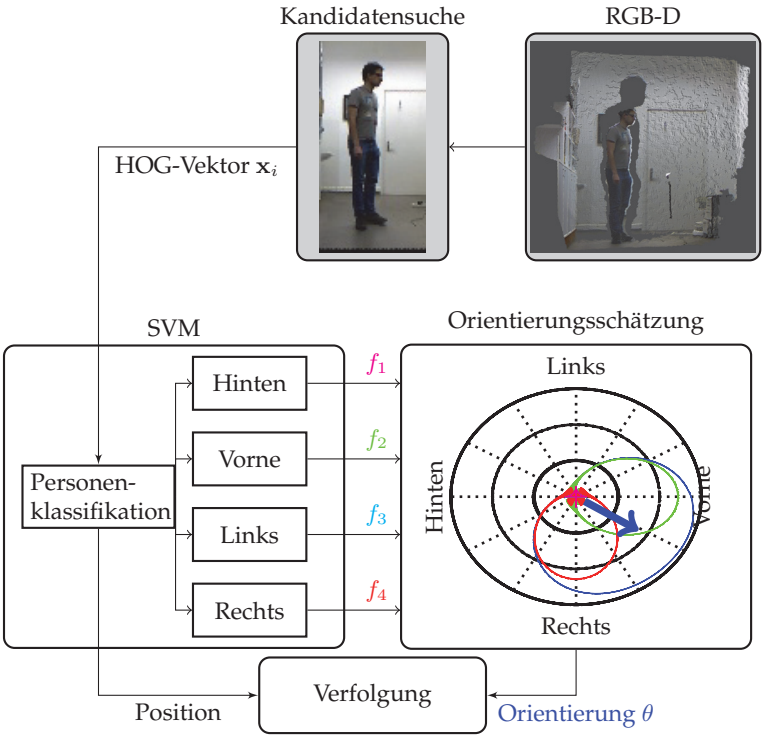


Abbildung 24.1: Ablaufdiagramm des vorgestellten Verfahrens zur richtungsabhängigen Detektion und Verfolgung von Fußgängern.

Vorauswahl durch maximale und minimale Gruppengröße wird hierdurch bereits eine effiziente Vorauswahl getroffen. Für die verbleibenden, typischerweise wenigen, Punktgruppen wird nun jeweils ein Detektionsfenster entsprechend der Abmessung der Punktgruppe skalierter, jedoch fester Größe im Farbbild betrachtet, für das der HOG-Merkmalsvektor berechnet wird. Diese Fenster können nun durch einen vorab trainierten SVM-Klassifikator auf Vorhandensein einer Person überprüft werden und erzeugen zusammen mit der Positionsinformation der entsprechenden Punktgruppe eine Detektion relativ zum Sensor, während Punktgruppen ohne Personen verworfen werden.

3 Orientierungsschätzung

Für die Schätzung der Orientierung kann der berechnete Merkmalsvektor weiterverwendet werden. In Anlehnung an [4] wird hier ebenfalls der Ansatz einer Mischung von „Experten"-Klassifikatoren gewählt. Allerdings unterscheidet sich der vorgestellte Ansatz deutlich durch die genaue Funktion und das Anlernen der Expertenmeinungen, da diese in diesem Beitrag nur für die Bestimmung der Orientierung und nicht zur Detektion von Personen eingesetzt werden und durch eine kontinuierliche Gewichtung der Trainingsdaten gezielter auf einzelne Richtungen angelernt werden. Die vier verwendeten Experten bestehen hier aus vier SVM-Klassifikatoren, die auf vier Hauptrichtungen k der Orientierungen hinten, vorne, links und rechts angelernt werden. Das Ergebnis jeder Klassifikation kann basierend auf den Trainingsdaten in eine Wahrscheinlichkeit für die Klassenzugehörigkeit $f_k(\mathbf{x}_i)$ des Merkmalsvektors $\mathbf{x}_i$ eines zu testenden Bildes transformiert werden [7]. Diese kann nun als Gewicht für die Bestimmung der Wahrscheinlichkeitsdichte der kontinuierlichen Orientierung θ verwendet werden:

$$p(\theta|\mathbf{x}_i) = \sum_{k=1}^{4} \frac{1}{4} f_k(\mathbf{x}_i)\, g_k(\theta)\,. \tag{24.1}$$

Hierbei wird wie auch in [4] jeder Experte zusammengesetzt aus der durch seine Wahrscheinlichkeit bestimmten Gewichtung f_k und sein Modell

$$g_k(\theta) = \mathcal{N}(\theta \,|\, \mu_k, \sigma^2)\,, \tag{24.2}$$

das durch eine Normalverteilung mit Mittelwert μ_k in die jeweilige der vier Hauptrichtungen und einer konstanten Standardabweichung σ beschrieben ist. Der Bereich mit dem höchsten Gewicht kann somit als beste Orientierung gewählt werden und den Positionsinformationen der Kandidatensuche zugeordnet werden. Eine beispielhafte Superposition der Experten mit bester Schätzung ist ebenfalls in Abbildung 24.1 zu sehen.

Das Anlernen der vier Klassifikatoren f_k spielt dabei eine entscheidende Rolle für die Bestimmung der Orientierung. Im Gegensatz zu [4] und der reinen Personendetektion aus Abschnitt 2 bestehen ne-

gative Trainingsbeispiele hier nicht aus Bildern ohne Fußgänger, sondern aus Bildern von Fußgängern anderer Hauptrichtungen, um eine bestmögliche Abgrenzung zu erreichen. Des Weiteren sollen zum einen beim Training Orientierungsunterschiede innerhalb einer Klasse von Hauptrichtungen berücksichtigt werden. Zum anderen soll auch eine stärkere Abgrenzung von Hauptrichtungen gegenüber der um 180° verschobenen, gegenüberliegenden Hauptrichtung im Vergleich zu den beiden verbleibenden, um 90° verschobenen Richtungen erreicht werden. Hierzu wird ein mit kontinuierlichen Orientierungen versehener Trainingsdatensatz verwendet und das normalverteilte Modell von g_k erneut aufgegriffen, das sowohl positive Trainingsbeispiele in einer gewichteten SVM bei größerer Abweichung von der idealen Hauptrichtung geringer gewichtet und Negativbeispiele je stärker gewichtet, desto mehr sie der gegenüberliegenden Hauptrichtung entsprechen.

4 Verfolgung

Die Literatur zur Verfolgung von Fußgängern geht zumeist von einem Modell konstanter Geschwindigkeit aus [2,3,8]. Auch wenn zusätzliche Erweiterungen nach sozialen menschlichen Verhalten [9] genutzt werden, bleiben weiterhin unvorhersehbare Richtungswechsel, Orientierungswechsel im Stand oder die Orientierung bei einer Initialisierung unberücksichtigt, da gerade diese Information bei der reinen Detektion fehlt und nur über zeitliche Bewegung erschlossen werden kann, was gerade im Fall von unsicherheitsbehafteten Messungen oder unbewegten Objekten zu Fehlern führen kann. Der hier vorgestellte Ansatz zur Verfolgung ist auf die Verwendung in einem Multi-Objekt-Ansatz zur Personenverfolgung nach [10] ausgelegt. Hier wird jedoch auf eine auf recheneffizientere Verwendung mittels des Unscented-Kalman-Filters [6] zurückgegriffen, das ebenfalls die Nutzung nichtlinearer Messmodelle erlaubt und die Funktionalität im Einobjektfall verdeutlicht. Im Zuge des Kalman-Update-Schritts werden neben dem unveränderten klassischen Positionsupdate nun die prädizierten Geschwindigkeiten $v_{\mathrm{x,präd}}$ und $v_{\mathrm{y,präd}}$ genutzt, um über dass Messmodell der Orientierung

h_{Ori} eine erwartete prädizierte Orientierung $\hat{\theta}$ zu erhalten:

$$\hat{\theta} = h_{\mathrm{Ori}}(v_{\mathrm{x,präd}}, v_{\mathrm{y,präd}}) = \arctan\left(\frac{v_{\mathrm{y,präd}}}{v_{\mathrm{x,präd}}}\right). \tag{24.3}$$

Die daraus erhaltene Abweichung zwischen $\hat{\theta}$ und der gemessenen Orientierung θ aus Abschnitt 3 kann nun als weiteres Element dem Kalman-Innovationsvektor zugeführt werden und zusammen mit einer vorgegebenen Unsicherheit in den weiteren Kalman-Gleichungen genutzt werden. Basierend auf den Unsicherheiten für Position und Orientierung kann somit eine Korrektur des prädizierten Zustands unter Berücksichtigung der zusätzlichen Information erfolgen.

5 Ergebnisse

Als Grundlage für Training und Evaluierung wird hier die KITTI-Datenbank [11] genutzt, da sie aus einer Vielzahl an unterschiedlichen Personen in unterschiedlichen Orientierungen in realistischer Umgebung mit Annotation der kontinuierlichen Orientierung besteht. Hieraus wurden alle unverdeckten und sich vollständig im Sensorsichtfeld befindenden Personen ausgeschnitten und auf die Fenstergröße des verwendeten Detektionsfensters von 128×64 Pixeln gebracht. Die Auftrennung der resultierenden 800 Bilder in Trainings- und Testdaten erfolgte zufällig je zur Hälfte. In allen Untersuchungen wurde wie in [4] der Parameter $\sigma = 45°$ gewählt.

5.1 Orientierung

Zur Beurteilung der Orientierungsschätzung wurden die auf den Testdaten geschätzten Orientierungen wieder in die vier Hauptrichtungen eingeteilt und mit den wahren Orientierungen der Bilder verglichen. Tabelle 24.1 zeigt die Konfusionsmatrix für die Zuordnung der vier Hauptrichtungen. Gerade die Zuordnung in rechte und linke Orientierung erreicht hier sehr gute Ergebnisse, wobei es bei Zuordnungen zur vorderen Richtung zu deutlich schlechteren Klassifikationsergebnissen kommt, was sich durch die im Vergleich zur oft charakteristischen Beinstellung von links/rechts wenig ausgeprägte Charakteristik

Tabelle 24.1: Konfusionsmatrix der Klassenzugehörigkeit bei einer Richtigklassifikationsrate von 85,2 %.

Wahr \ Schätzung	Hinten	Vorne	Links	Rechts
Hinten	0.88	0.04	0.04	0.03
Vorne	0.08	0.64	0.13	0.15
Links	0.05	0.02	0.93	0.01
Rechts	0.02	0.06	0	0.92

von Orientierungen wie vorne und hinten begründen lässt. Die durchschnittliche Richtigklassifikationsrate liegt bei 85,2 %, was eine deutliche Verbesserung gegenüber dem Ansatz aus [4] darstellt, der mit dem hier gewählten Datensatz eine Rate von 70 % erreichte, wobei im ursprünglichen Ansatz zusätzlich Konturinformationen verwendet wurden, die hier nicht berücksichtigt wurden.

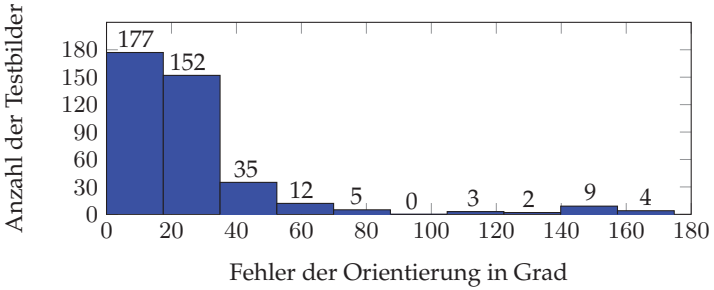


Abbildung 24.2: Häufigkeiten der Beträge der absoluten Orientierungsfehler unter den getesteten Bildern.

In Abbildung 24.2 sind die Beträge der absoluten Fehler der kontinuierlichen Orientierungsschätzung als Histogramm angegeben. Es ist gut zu sehen, dass – wie auch an den Klassifikationsergebnissen zu erkennen – eine Fehlschätzung um 90° seltener vorkommt als eine Verwechslung der gegensätzlichen Seite. Der mittlere absolute Winkelfehler lässt sich zu 29,8° bestimmen und zeigt, dass die exakte Ausrichtung trotz hoher Richtigklassifikationsraten für die verschiedenen Ausrichtungen deutlich mit einer zu berücksichtigenden Unsicherheit geschätzt wird. Diese wird jedoch unabhängig von der Position der Person gewonnen

und stellt somit eine sehr gute Ergänzung zur Schätzung der Orientierung alleine aus Geschwindigkeitsvektoren dar.

5.2 Verfolgung

Zur Validierung der Ergebnisse für die Fusion von zeitlichen Positionsinformationen mit den Orientierungen wurde aus Gründen der Anschaulichkeit ein selbst erzeugtes Ein-Personen-Szenario mit häufigen Richtungswechseln gewählt. Hierfür wurden manuell die wahren Orientierungen (schwarz) annotiert. Abbildung 24.3 zeigt die Ergebnisse für die Unscented-Kalman-Filterung mit (rot) und ohne (blau) Orientierungsinformationen zusammen mit Farbbildern zu ausgewählten Zeitschritten. Während beide Filter aufgrund des Modells konstanter Geschwindigkeit ein eher träges Verhalten aufzeigen, ist deutlich zu sehen,

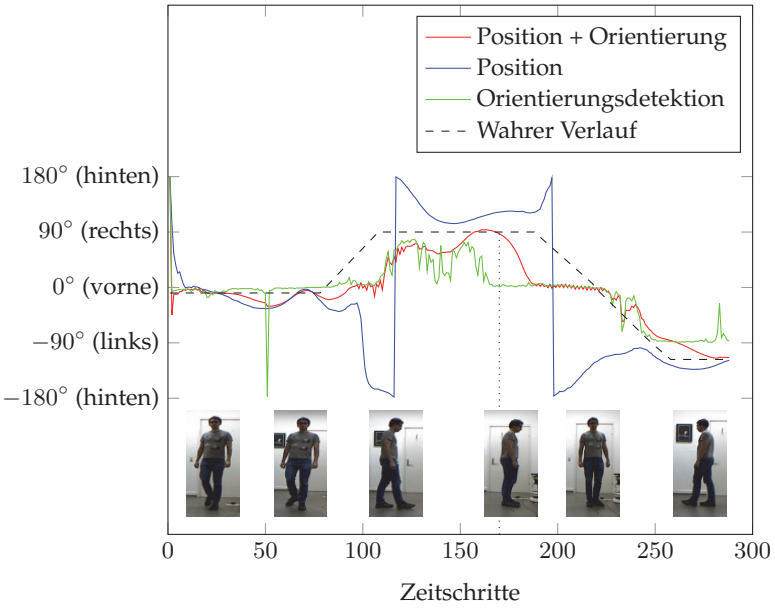


Abbildung 24.3: Verfolgung des Orientierungswinkels mit und ohne Orientierungsschätzung bei gegebenem annotierten Verlauf.

dass die Orientierungsinformation gerade die Schätzung der Drehrichtung korrigiert, die bei der Filterung ohne Orientierungsinformation nur auf Bewegung basiert und somit an unbewegten Richtungswechseln zu Fehlern führen kann. Diese Fehler werden erst durch eine auf die Richtungsänderung folgende Bewegung korrigiert. Hingegen wird die – wie beispielsweise um Zeitschritt 170 (gepunktete Linie) – fehlerhafte Orientierungsdetektion (grün) durch die vorhergehende eindeutige Bewegungsrichtung ausgeglichen.

6 Zusammenfassung

Im Beitrag wurde ein Rahmenwerk für die Fusion von Farb- und Tiefenbildfolgen zum Zwecke der Personenverfolgung vorgestellt. Neben der geschätzten Position der zu detektierenden Personen, die aus den Tiefendaten resultiert, verwendet der Ansatz als Zusatzinformation eine aus den Farbbildern gewonnene Schätzung ihrer kontinuierlichen Orientierung. Die Orientierung wird nach der Personendetektion durch Verwendung eines HOG-Merkmalsvektors bestimmt, der durch vier Expertenklassifikatoren für die vier Hauptrichtungen ausgewertet wird. Dieses Zusatzwissen führt zu Verbesserungen der Filterung in Fällen, in denen die Annahmen des üblicherweise verwendeten Modells konstanter Geschwindigkeit verletzt werden. Für zukünftige Untersuchungen ist geplant, weitere Merkmale für die gezieltere Unterscheidung ähnlicher Fälle wie vorne/hinten zu verwenden und auch die durch die Superposition der Experten ausgedrückte Unsicherheit adaptiv zu nutzen. Ebenso stehen die Evaluierung auf einer größeren Datenbank und im Multi-Objekt-Fall noch aus.

Literatur

1. N. Dalal und B. Triggs, „Histograms of oriented gradients for human detection“, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, June 2005, S. 886–893 vol. 1.
2. L. Spinello und K. Arras, „People detection in RGB-D data“, in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2011, S. 3838–3843.
3. M. Munaro, F. Basso und E. Menegatti, „Tracking people within groups with RGB-D data“, in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2012, S. 2101–2107.
4. M. Enzweiler und D. Gavrila, „Integrated pedestrian classification and orientation estimation“, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2010, S. 982–989.
5. A. Pérez Grassi, V. Frolov und F. Puente León, „Information fusion to detect and classify pedestrians using invariant features“, *Information Fusion*, Vol. 12, Nr. 4, S. 284 – 292, 2011, special Issue on Information Fusion for Cognitive Automobiles.
6. S. Julier und J. Uhlmann, „Unscented filtering and nonlinear estimation“, *Proceedings of the IEEE*, Vol. 92, Nr. 3, S. 401–422, Mar 2004.
7. J. C. Platt, „Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods“, in *Advances in large margin classifiers*. MIT Press, 1999, S. 61–74.
8. C. Otto, W. Gerber, F. Puente León und J. Wirnitzer, „A joint integrated probabilistic data association filter for pedestrian tracking across blind regions using monocular camera and radar“, in *IEEE Intelligent Vehicles Symposium (IV)*, 2012, S. 636–641.
9. D. Helbing und P. Molnar, „Social force model for pedestrian dynamics“, *Physical Review*, Vol. 51, S. 4282–4286, 1995.
10. J. Pallauf und F. Puente León, „State-dependent and distributed pedestrian tracking using the (C)PHD filter“, in *IEEE International Instrumentation and Measurement Technology Conference*, Montevideo, Uruguay, 12-15 May 2014.
11. A. Geiger, P. Lenz und R. Urtasun, „Are we ready for autonomous driving? The KITTI vision benchmark suite“, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2012, S. 3354–3361.

Streifenprojektionsgenauigkeit mit Kinect-Rate – 3D-Sensorik für schnelle, dichte und genaue Formvermessung

M. Schaffer und M. Große

EnShape GmbH
c/o Institut für Angewandte Optik
Fröbelstieg 1, 07743 Jena

Zusammenfassung Dieser Beitrag stellt ein Verfahren zur optischen 3D-Vermessung von Oberflächenformen auf Basis der Stereophotogrammetrie und Projektion von statistischen Mustern vor. Der Vorteil im Vergleich zu konventionellen Streifenprojektionssystemen besteht darin, dass die Muster sehr schnell projiziert und damit sehr kurze Messzeiten realisiert werden können. Des Weiteren wird die Reduzierung der Auswertezeit vorgestellt, so dass das Verfahren in der Lage ist, sowohl schnell zu messen, als auch schnell 3D-Daten bereit zu stellen. Ein Vergleich mit einem etablierten Streifenlichtsensor zeigt, dass die hohe Aufnahmegeschwindigkeit nicht zulasten der Messpräzision geht und insbesondere die Projektion statistischer Muster gleiche Messpräzisionen wie die Streifenprojektion erreichen kann.

1 Einleitung

Optische Verfahren auf Basis der Musterprojektion werden zunehmend für die dreidimensionale Erfassung von Oberflächen eingesetzt. Sie zeichnen sich insbesondere durch ihren geringen Zeitbedarf aus, in dem sie viele 3D-Punkte des Messobjekts erfassen können. Im Forschungsumfeld zeichnen sich momentan zwei Richtungen der flächigen 3D-Sensorik ab. Zum einen sehr schnelle Verfahren, die 3D-Daten als Stream mit Raten von bis zu 30 Hz zur Verfügung stellen – bspw. Time of Flight oder Microsoft Kinect. Zum anderen sehr genaue Verfahren, die Hunderttausende sehr genau lokalisierte 3D-Punkte pro Einzelmessungen generieren – bspw. Stereophotogrammetrie mit strukturierter

Beleuchtung, Streifenprojektion. Während erstere vergleichsweise ungenau sind, sind letztere eher langsam (≤ 1 Hz).

Ein Forschungsbedarf besteht demnach in der Entwicklung von Verfahren, die, sowohl schnell als auch hochgenau, flächig viele 3D-Punkte erfassen können. Für stereophotogrammetrische Verfahren mit strukturierter Beleuchtung besteht die Herausforderung darin, die sequentielle Musterprojektion auf die Rate der Aufnahme der Kameras zu erhöhen, d.h. bei Kameraaufnahmeraten von 1000 Hz, Muster (bspw. Streifenmuster) mit dieser Rate zu projizieren.

Dies ist mit aktuell verfügbaren Projektoren – üblicherweise DMD-Projektoren – nicht möglich [1, 2]. Um den Flaschenhals der Musterprojektion zu umgehen, wurden deshalb im Forschungsumfeld verschiedene Verfahren zur hochfrequenten Strukturierung der Messoberfläche entwickelt. Die Defokussierung binärer Streifenmuster wurde in [3] umgesetzt und nutzt die hohen Schaltraten von DMDs für binäre Musterbilder. Die Defokussierung erzeugt keine exakten 1+Cos-Streifenmuster, so dass periodische Artefakte in den 3D-Rekonstruktionen sichtbar sind [3]. Auch die Anzeige von Graycodes und damit Verstetigung der Messdaten wird durch die Defokussierung erschwert [4]. Neue Projektortypen, die auf Einzelprojektoren pro Muster [5], Multi-LED-Projektion [6] oder Temporal Dithering [7] beruhen, wurden entwickelt, die schnell Streifenmuster projizieren bzw. in ihrer Phase verschieben können.

Ehemalige Mitarbeiter der Arbeitsgruppe 3D-Vermessung und Gründer des Start-Ups EnShape GmbH der Friedrich-Schiller-Universität Jena haben ein Verfahren entwickelt, welches auf Basis der Projektion statistischer Muster eine sehr hohe Musterprojektionsrate erreicht und gleichzeitig zu keiner Reduzierung der Messgenauigkeit bei der 3D-Rekonstruktion führt.

2 3D-Aufnahmen mit kurzer Messzeit

Der schematische Aufbau eines 3D-Sensors auf Stereophotogrammetriebasis ist in Abbildung 25.1 gezeigt. Zwei Kameras $\{1, 2\}$ betrachten das Messobjekt aus unterschiedlichen Perspektiven. Während der Messung werden mehrere Bilder (N), jeweils zum Zeitpunkt t, des Messobjekts aufgenommen. Zeitgleich wird zu jedem der Kamerabilder ein

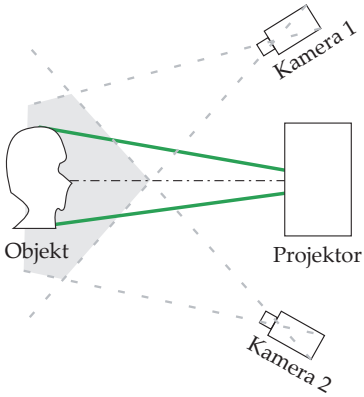


Abbildung 25.1: Aufbau des Stereophotogrammetriesensors mit High-Speed Kameras und Musterprojektor.

statistisches Muster projiziert, so dass eine optimale Strukturierung der Objektoberfläche erreicht wird. Diese Strukturierung ist notwendig, da die Suche homologer Punkte im Anschluss an die Messung, insbesondere für nicht-texturierte Objekte, durch das Aufbringen künstlicher Lichtstrukturen signifikant verbessert wird.

3 Punktwolkenrekonstruktion

Im Anschluss an die Bilddatenaufnahme und Musterprojektion erfolgt die Rekonstruktion der 3D-Punkte aus den Bilddaten. Jedem Pixel p_i^1, p_j^2 der jeweiligen Ansicht kann eine Sequenz von N Grauwerten zugeordnet werden, die der Helligkeit entsprechen, die zum Zeitpunkt t durch die Musterprojektion auf dem assoziierten Objektbereich erzeugt wurde. Vergleicht man die Sequenzen der Pixel von Kamera 1 mit denen von Kamera 2, können korrespondierende oder auch homologe Punkte bestimmt werden (siehe Schema in Abbildung 25.2), die einen ähnlichen Grauwertverlauf aufzeigen und dem selben Objektpunkt zugehörig sind.

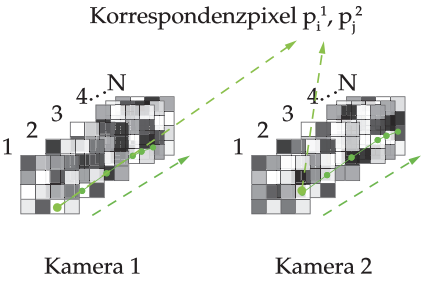


Abbildung 25.2: Aufbau des Stereophotogrammetriesensors mit High-Speed Kameras und Musterprojektor.

Unter Zuhilfenahme von Gleichung (25.1) kann fur jedes Pixelpaar $\left(p_i^1, p_j^2\right)$ der Korrelationswert ρ bestimmt werden.

$$\rho\left(p_i^1, p_j^2\right) = \frac{\sum_{t=1}^{N}\left(g_{p_i^1,t} - \overline{g}_{p_i^1}\right)\cdot\left(g_{p_j^2,t} - \overline{g}_{p_j^2}\right)}{\sqrt{\sum_{t=1}^{N}\left(g_{p_i^1,t} - \overline{g}_{p_i^1}\right)^2}\cdot\sqrt{\sum_{t=1}^{N}\left(g_{p_j^2,t} - \overline{g}_{p_j^2}\right)^2}} \tag{25.1}$$

Dieser wird als quantitatives Maß verwendet, um die Ähnlichkeit der Grauwertverläufe zu ermitteln. Der korrespondierende Pixel p_j^2 zu p_i^1 mit dem höchsten Korrelationswert wird als homolog gekennzeichnet und für die anschließende Triangulation verwendet. Durch Verwendung eines zusätzlichen Schwellwerts ρ_{Min} können schwach korrelierende Bildpunkte, die üblicherweise in Ausreißerpunkte resultieren, unterdrückt werden.

Nach Berechnung aller Bildpunktkorrespondenzen werden diese trianguliert. Die vor der Messung erfolgte Kalibrierung von inneren und äußeren Parametern wird dabei genutzt, um für jedes Korrespondenzpaar einen 3D-Punkt $P = \{x, y, z\}$ zu berechnen.

4 Experimentelle Umsetzung

Infolge der sehr hohen Projektionsrate von experimentell gezeigten 20.000 Hz [8] können sehr kurze Messzeiten realisiert werden. Diese werden im Folgenden in Unterabschnitt 4.1 und 4.2 in zwei getrennten Szenarien genutzt und beschrieben.

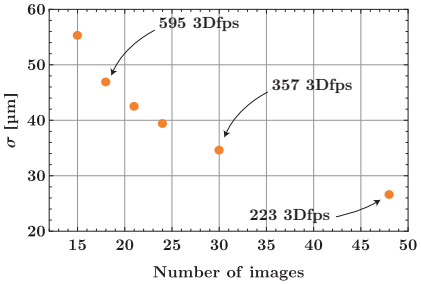


Abbildung 25.3: Messpräzision σ in Abhängigkeit der Sequenzlänge N.

4.1 Offline-Processing

In einer experimentellen Umsetzung des Aufbaus in Abbildung 25.1 wurden zwei Kameras vom Typ PCO Dimax HD verwendet. Die Aufnahmerate bei einer Auflösung von 720×480 betrug $10.700\,\mathrm{Hz}$. Der Basisabstand der Kameras belief sich auf $0,25\,\mathrm{m}$ und der Objektabstand zur Basis glich $0,7\,\mathrm{m}$. Die Bilddaten wurden in der Kamera zwischengespeichert und nach Übertragung auf den Computer wie oben beschrieben ausgewertet. Um Aussagen über die Messgenauigkeit zu treffen, wurde die Messpräzision σ als Wiederholstandardabweichung eines 3D-Punktes in den rekonstruierten Punktwolken bestimmt. Dazu wurde eine zertifizierte Ebene (Kalibrierstandard, Planarität $3,4\,\mu\mathrm{m}$, Rauheit $\leq 0,2\,\mu\mathrm{m}$) im Messvolumen von $22,5 \times 15 \times 10\,\mathrm{cm}^3$ platziert, sodass nach Abzug eines Ebenenfits das Rauschen um die Ebene als Wert σ bestimmt werden konnte. Das Rauschen der 3D-Punkte σ für unterschiedliche N ist in Abbildung 25.3 als Maß für die Messpräzision gezeigt.

Für 10.700 Bildpaare pro Sekunde und einer Sequenzlänge N von bspw. 30 Bildern in der Rekonstruktion erhält man eine Messpräzision von $\sigma = 35\,\mu\mathrm{m}$ bei einer Messrate von beachtlichen 357 3D-Messungen pro Sekunde. Jede 3D-Messung beinhaltete dabei maximal 345.600 3D-Punkte mit der Wiederholstandardabweichung σ pro 3D-Punkt. Dies entspricht einer relativen Messpräzision von $1,2 \cdot 10^{-4}$ – Messpräzision relativ zu Diagonale des Messvolumens – bei einer Messzeit von lediglich $2,8\,\mathrm{ms}$.

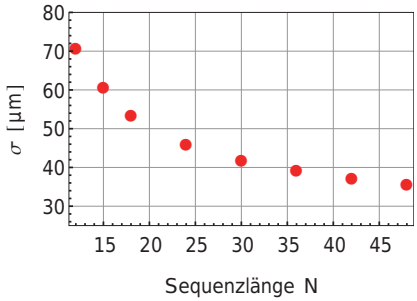


Abbildung 25.4: Messpräzision σ in Abhängigkeit der Sequenzlänge N.

4.2 Online-Processing

Für Anwendungen, wie die Qualitätskontrolle am Fließband oder Pick & Place-Aufgaben, ist nicht nur eine hohe Messgenauigkeit und damit Messpräzision relevant, sondern ebenso eine kurze Rekonstruktionszeit, sodass bestenfalls unmittelbar nach der Messung die Punktwolke und damit die dreidimensionale Repräsentation der Oberfläche zur Inspektion und Interpretation vorliegt. Dadurch ist es notwendig neben der Messzeit auch die Auswertezeit signifikant zu reduzieren.

In einem weiteren Aufbau wurde deshalb ebenfalls eine schnelle Auswertung implementiert. Dieser ist auch im Sensorprototyp der EnShape GmbH realisiert. Der Basisabstand der Kameras beträgt dabei $0,4\,\mathrm{m}$ und der Abstand zum Messobjekt beläuft sich auf $1\,\mathrm{m}$, wobei das beleuchtete und erfasste Messfeld circa der Größe A4 mit Abmessungen von $30 \times 22,5\,\mathrm{cm}^2$ entspricht.

Im Anschluss an die Messzeit von $30\,\mathrm{ms}$, für den Fall $N = 30$, wird die Rekonstruktion in weiteren $60\,\mathrm{ms}$ durchgeführt, so dass nach circa $90\,\mathrm{ms}$ die 3D-Punktwolke als Repräsentation der Messoberfläche vorliegt. Durch eine Überlappung von Messung und Auswertung kann eine Messrate von $\geq 30\,\mathrm{Hz}$ und damit eine Zykluszeit von $\leq 33\,\mathrm{ms}$ erreicht werden. Ebenfalls wurde die Messpräzision an der zertifizierten Referenzebene überprüft und das 3D-Punktrauschen bestimmt. Entsprechend Abbildung 25.4 kann bei dieser kurzen Zykluszeit ein Messrauschen von $\sigma = 42\,\mu\mathrm{m}$ erzielt werden, sodass die 3D-Vermessung in dieser Implikation sowohl als schnell als auch präzise

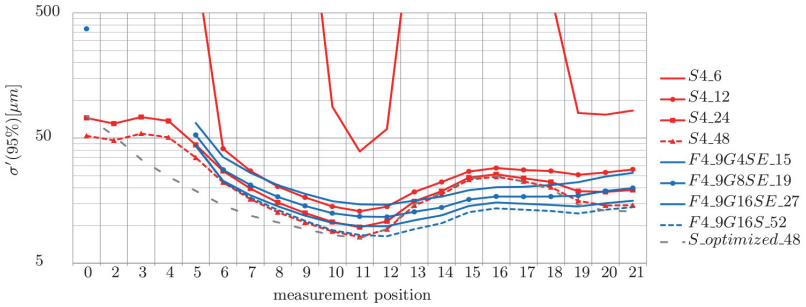


Abbildung 25.5: Messpräzision σ in Abhängigkeit des Messobjektabstandes. Streifenprojektion in blau mit variierender Gesamtbildanzahl und Projektion statistischer Muster in rot mit variierender Gesamtbildanzahl (Sequenzlänge N). 5% der entferntesten 3D-Punkte wurden entfernt.

bezeichnet werden kann. Die relative Messpräzision – zum Vergleich – beläuft sich auf $1,5 \cdot 10^{-4}$.

5 Vergleich zur Streifenprojektion

Momentan ist die Standardtechnologie im Bereich der stereophotogrammetrischen, hochgenau optischen Messverfahren makroskopischer Messfelder die Streifenprojektion. Dabei wird ein $1 + \cos$-Muster vom Projektor auf das Objekt projiziert und zwischen jeder Bildpaaraufnahme um einen definierten Phasenwert verschoben. Da dieses Verfahren weit verbreitet ist und als eines der genausten gilt, wurde die o.g. statistische Musterprojektion mit der Projektion phasengeschobener Streifenmuster eingehend verglichen [9]. Der Vergleich wurde mit einem Stereophotogrammetrie-System des Fraunhofer IOF, Jena, durchgeführt, wobei die um $25\,\mathrm{cm}$ räumlich getrennten Kameras eine Auflösung von $2452 \times 2054\,\mathrm{px}$ auszeichnete und das Messfeld eine Diagonale von $18\,\mathrm{cm}$ aufwies. Durch die reine Veränderung der Projektionsstrukturen innerhalb des DMD-Projektors konnte ein Einfluss der Kameras, Objektive, Kalibrierung und Messfeldgröße auf die 3D-Vermessung ausgeschlossen werden. Die Ergebnisse sind in Abbildung 25.5 gezeigt.

Es ist erkennbar, dass analog zu den Abbildungen 25.3 und 25.4 mit steigender Anzahl an Mustern das Rauschen der 3D-Punkte um die rekonstruierte Messebene sinkt. Dies gilt sowohl für die Anzahl der Phasenschritte im Falle der Streifenprojektion als auch für die Anzahl der statistischen Muster. Je mehr Muster verwendet werden, desto präziser kann der 3D-Punkt rekonstruiert werden. Des Weiteren zeigt sich für Objektpositionen außerhalb des Schärfentiefebereichs von Projektor und Kameras, dass das 3D-Punktrauschen stark ansteigt. Für eine Sequenzlänge von $N = 6$ kommt es zu vielen Ausreißern, insbesondere im Bereich unscharfer Musterprojektion, da die Bildpunktzuordnung mittels zeitlicher Korrelation signifikant beeinträchtigt wird. Hingegen kann für eine Sequenzlänge von $N = 48$ eine Wert σ von $8,6\,\mu\mathrm{m}$ mit statistischen Mustern erreicht werden, der einer relativen Messpräzision von $4,8 \cdot 10^{-5}$ entspricht. Durch die signifikant höhere Auflösung dieses Stereophotogrammetrieaufbaus kann dieser geringe Wert im Vergleich zu den o.g. erklärt werden.

6 Zusammenfassung

Durch den Einsatz statistischer Muster ist es möglich sehr hohe Projektionsraten zu realisieren. In Experimenten konnten bis zu $20.000\,\mathrm{Hz}$ erreicht werden und Kameras mit einer Aufnahmerate von $10.700\,\mathrm{Hz}$ für sehr schnelle 3D-Vermessungen mit Raten von mehreren Hundert Hertz verwendet werden. Gleichzeitig wurde die Messpräzision nicht reduziert, sodass jeder 3D-Punkt eine Wiederholstandardabweichung in seiner Lage von $35\,\mu\mathrm{m}$ aufwies.

Das Verfahren ist in der Auswertung flexibel, so dass zum Einen nachträglich die Anzahl der Muster reduziert werden kann und damit Messpräzision und Messzeit gewählt werden können. Zum Anderen kann die Auswertung der Bilddaten mittels zeitlicher Korrelation stark beschleunigt werden, so dass Messung und Auswertung weniger als $100\,\mathrm{ms}$ benötigen. Durch Überlappung von Messung und Auswertung kann sogar eine Rate von $30\,\mathrm{Hz}$ erreicht werden. Solche Messraten werden üblicherweise nur von Spielemarktsensoren, wie der Microsoft Kinect erreicht, die jedoch eine signifikant reduzierte Messpräzision aufweisen und üblicherweise systematische Messfehler mit sich bringen.

In Vergleichsmessungen mit einem etablierten Streifenprojektionssys-

tem konnte gezeigt werden, dass die Wahl der Musterstruktur (Streifenmuster oder statistische Muster) keinen Einfluss auf die erreichbare Messpräzision hat, sodass trotz der kurzen Messzeit und Auswertezeit eine hohe Messpräzision, wie üblicherweise nur von Streifenprojektionssystemen erwartet, erreicht wird und damit Objekte in kurzer Zeit hochgenau dreidimensional vermessen werden können.

Literatur

1. S. Zhang, „Recent progresses on real-time 3d shape measurement using digital fringe projection techniques“, *Optics and Lasers in Engineering*, Vol. 48, S. 149–158, 2010.
2. J. Salvi, S. Fernandez, T. Pribanic und X. Llado, „A state of the art in structured light patterns for surface profilometry“, *Pattern Recognition*, Vol. 43, Nr. 8, S. 2666 – 2680, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003132031000124X
3. Y. Wang und S. Zhang, „Superfast multifrequency phase-shifting technique with optimal pulse width modulation“, *Opt. Express*, Vol. 19, Nr. 6, S. 5149–5155, Mar 2011. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-19-6-5149
4. ——, „Three-dimensional shape measurement with binary dithered patterns“, *Appl. Opt.*, Vol. 51, Nr. 27, S. 6631–6636, Sep 2012. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-51-27-6631
5. S. Heist, M. Sieler, A. Breitbarth, P. KÃ¼hmstedt und G. Notni, „High-speed 3d shape measurement using array projection“, S. 878 815–878 815–11, 2013. [Online]. Available: http://dx.doi.org/10.1117/12.2020539
6. S. Zwick, S. Heist, R. Steinkopf, S. Huber, S. Krause, C. Braeuer-Burchardt, P. Kuehmstedt und G. Notni, „3D phase-shifting fringe projection system on the basis of a tailored free-form mirror“, *APPLIED OPTICS*, Vol. 52, Nr. 14, S. 3134–3146, MAY 10 2013.
7. P. Wissmann, R. Schmitt und F. Forster, „Fast and accurate 3d scanning using coded phase shifting and high speed pattern projection“, in *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2011 International Conference on*, May 2011, S. 108–115.
8. M. Grosse, M. Schaffer, B. Harendt und R. Kowarschik, „Fast data acquisition for three-dimensional shape measurement using fixed-pattern projection and temporal coding“, *Optical Engineering*, Vol. 50, Nr. 10, S. 100503, 2011. [Online]. Available: http://link.aip.org/link/?JOE/50/100503/1

9. P. Lutzke, M. Schaffer, P. Kühmstedt, R. Kowarschik und G. Notni, „Experimental comparison of phase-shifting fringe projection and statistical pattern projection for active triangulation systems“, *Proceedings of the SPIE*, Vol. 8788, S. 878 813–878 813–7, 2013. [Online]. Available: +http://dx.doi.org/10.1117/12.2020910

A different approach to multi-period phase shift

Thomas Dunker and Sebastian Luther

Fraunhofer Institute for Factory Operation and Automation IFF
Sandtorstraße 22, D-39106 Magdeburg

Abstract Phase shift measurements with different wavelengths allow unwrapping the period numbers in a certain range of unambiguity. This paper proposes a geometric interpretation of this method in order to understand the influence of measurement noise. An application of this approach is to find three wavelengths with some technical constraints, which e.g. for given resolution requirement tolerate the highest standard deviation of the measurement noise.

1 Introduction

Starting point of our reflections is the method for reconstructing projector coordinates using phase shifts of sinusoidal patterns of different wavelengths reported in [1]. This method gives an efficient algorithm for computing period numbers from phase measurements.

This was the method of choice for an infrared fringe projection system developed by AiMESS Services GmbH, Burg, Germany, which projects patterns of three wavelengths, see [2].

One important point was to understand, which wavelength combination for the gratings is most tolerant against fabrication tolerances of the gratings and noise in the phase measurement.

The authors of [3] analyze similar questions. Given wavelengths and maximal measurement noise, they determine the maximal range, in which period numbers can be computed unambiguously from phase measurements.

2 Notation

We project k periodic fringe patterns with period lengths $\overline{\lambda} \geq \lambda_1 > \ldots > \lambda_k \geq \underline{\lambda} > 0$. For each pattern we define a coordinate ξ_i, which can be decomposed into period number $\eta_i = \lfloor \xi_i \rfloor \in \mathbb{Z}$ and phase $\varphi_i = \xi_i \bmod 1 \in [0, 1)$. There is a common projector coordinate with $\zeta = \lambda_i(\xi_i - \xi_i^0)$ for all $i = 1, \ldots, k$. In the sequel we will assume all offsets $\xi_i^0 = 0$.

We can interpret the pattern coordinates $\boldsymbol{\xi}(\zeta) = (\xi_1, \ldots, \xi_k)^T$ as a line segment

$$\Xi = \{\boldsymbol{\xi}(\zeta) \in \mathbb{R}^k : \zeta \in [\underline{\zeta}, \overline{\zeta}]\}.$$

We denote by $\boldsymbol{\tau}^0 = (\lambda_1^{-1}, \ldots, \lambda_k^{-1})^T$ a tangent vector of $\boldsymbol{\xi}(\zeta)$ and let $\boldsymbol{\tau} = \boldsymbol{\tau}^0 / \|\boldsymbol{\tau}^0\|$ be normalized. A possible combination of period numbers $\boldsymbol{\eta} \in \mathbb{Z}^k$ can be associated with the cube $\boldsymbol{\eta} + [0, 1)^k$, which intersects the line segment.

As both camera and projector add some noise, we model our phase measurement as $\tilde{\boldsymbol{\varphi}} = (\boldsymbol{\xi} + \boldsymbol{\varepsilon}) \bmod 1 \in [0, 1)^k$, where $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{C})$ is normally distributed with covariance $\boldsymbol{C}$. We assume that covariance $\boldsymbol{C}$ has full rank. Then it defines a scalar product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{C}} = \boldsymbol{x}^T \boldsymbol{C}^{-1} \boldsymbol{y}$ on $\mathbb{R}^k$. Using this we define a projection

$$\boldsymbol{P}_{\parallel} = \frac{\boldsymbol{\tau}\boldsymbol{\tau}^T \boldsymbol{C}^{-1}}{\boldsymbol{\tau}^T \boldsymbol{C}^{-1} \boldsymbol{\tau}}$$

onto the line. One can verify that $\hat{\boldsymbol{\xi}} = \boldsymbol{P}_{\parallel} \tilde{\boldsymbol{\xi}}$ is an unbiased estimate and the covariance $(\boldsymbol{\tau}^T \boldsymbol{C} \boldsymbol{\tau}) \boldsymbol{\tau}\boldsymbol{\tau}^T$ of $\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}$ is identical to the one of $\boldsymbol{\tau}\boldsymbol{\tau}^T \boldsymbol{\varepsilon}$ - the noise in direction of the line. Remark that e.g. for $\boldsymbol{C} = \boldsymbol{I}$ the variances of the single components $\tau_i^2 < 1$ are smaller than the variance of the initial noise.

For the above we still need the integer part $\lfloor \tilde{\boldsymbol{\xi}} \rfloor$. We use the notation $\hat{\boldsymbol{\eta}} = \lfloor \tilde{\boldsymbol{\xi}} \rfloor$ for the estimate of the period numbers. Let us denote by H the set of all candidates $\boldsymbol{\eta}$ and let $\boldsymbol{P}_{\perp} = \boldsymbol{I} - \boldsymbol{P}_{\parallel}$. Then a best estimate for the period numbers $\hat{\boldsymbol{\eta}}(\tilde{\boldsymbol{\varphi}}) \in H$ is defined by

$$\begin{aligned} \|\boldsymbol{P}_{\perp}(\hat{\boldsymbol{\eta}}(\tilde{\boldsymbol{\varphi}}) + \tilde{\boldsymbol{\varphi}})\|_{\boldsymbol{C}} &\leq \|\boldsymbol{P}_{\perp}(\boldsymbol{\eta} + \tilde{\boldsymbol{\varphi}})\|_{\boldsymbol{C}} \\ &= \|\boldsymbol{D}^{-1/2} \boldsymbol{U}^T \boldsymbol{P}_{\perp}(\tilde{\boldsymbol{\varphi}} - (-\boldsymbol{\eta}))\| \end{aligned}$$

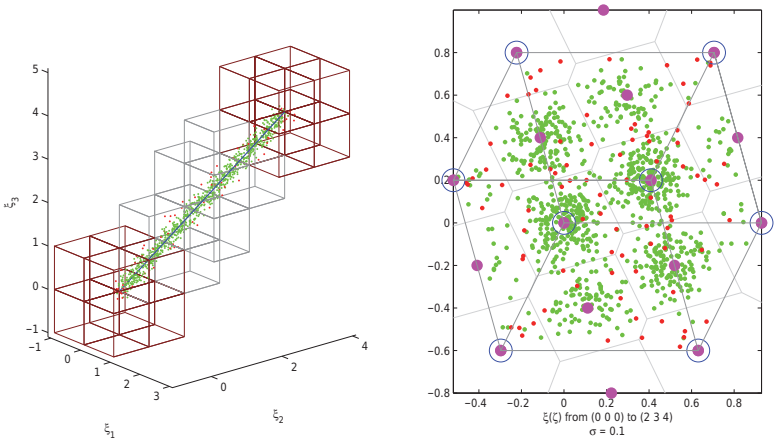


Figure 26.1: On the left, sketch of the line segment Ξ with a wavelength ration of $1 : 2/3 : 1/2$ and randomly disturbed points, which yield the disturbed phase measurements $\tilde{\boldsymbol{\varphi}} = (\boldsymbol{\xi} + \boldsymbol{\varepsilon}) \bmod 1$ with $\boldsymbol{C} = \sigma^2 \boldsymbol{I}$. The possible combinations of period numbers $\boldsymbol{\eta}$ for all points are illustrated by cubes. Because of the noise there are although cubes not intersecting the line segment. Red cubes differ by $\boldsymbol{\xi}(\operatorname{lcm}(\lambda_1, \ldots, \lambda_k))$ and are mapped to the same points in H_{ref}. On the right – in the 2-dimensional image of the projection $\boldsymbol{P}_\perp(-H)$ and $\boldsymbol{P}_\perp \tilde{\boldsymbol{\varphi}}$ – they are marked by blue circles. Using the voronoi cells for assigning the unknown period numbers to the phase measurements, the period numbers of the green points are unwrapped correctly while the red ones failed.

for all $\boldsymbol{\eta} \in H$, where $\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T$ is the eigenvalue decomposition of the covariance $\boldsymbol{C}$. The $(k-1)$ dimensional image of $\boldsymbol{D}^{-1/2}\boldsymbol{U}^T\boldsymbol{P}_\perp$ is orthogonal on the null space of $\boldsymbol{P}_\perp^T\boldsymbol{U}\boldsymbol{D}^{-1/2}$, which is the span of $\boldsymbol{D}^{-1/2}\boldsymbol{U}^T\boldsymbol{\tau}$. Thus for given $\boldsymbol{D}^{-1/2}\boldsymbol{U}^T\boldsymbol{P}_\perp\tilde{\boldsymbol{\varphi}}$ we need to find the closest point from the set $H_{\text{ref}} = \boldsymbol{D}^{-1/2}\boldsymbol{U}^T\boldsymbol{P}_\perp(-H)$. i.e. H_{ref} partitions the $(k-1)$ subspace in voronoi cells. The closer two points in H_{ref} to each other the higher is the probability that a wrong $\boldsymbol{\eta}$ might be assigned.

If the wavelengths $\lambda_1, \ldots, \lambda_k$ have rational ratios then there exists a least common multiple $\operatorname{lcm}(\lambda_1, \ldots, \lambda_k)$, which limits the range of unambiguity by $\boldsymbol{\varphi}(\zeta) = \boldsymbol{\varphi}(\zeta + \operatorname{lcm}(\lambda_1, \ldots, \lambda_k))$. In this case, if two points in H differ by $\boldsymbol{\xi}(\operatorname{lcm}(\lambda_1, \ldots, \lambda_k))$, they are mapped to the same point in H_{ref}.

In figure 26.1 we have pairs of cubes $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2$, which differ by exactly $\boldsymbol{\xi}(\mathrm{lcm}(\lambda_1, \ldots, \lambda_k))$ yielding $\|\boldsymbol{P}_\perp(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2)\|_{\boldsymbol{C}} = 0$. We can observe in addition that for all other pairs of cubes of figure 26.1 it holds $|\boldsymbol{\tau}^T(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2)| < \|\boldsymbol{\xi}(\mathrm{lcm}(\lambda_1, \ldots, \lambda_k))\|$. That is why, we propose the following kind of set

$$H(\Xi, \boldsymbol{C}) = \{\boldsymbol{\eta} \in \mathbb{Z}^k : t_1 \leq \boldsymbol{\tau}^T \boldsymbol{\eta} \leq t_2\} \tag{26.1}$$

as candidate set for the unwrapping, where $t_1 < t_2$ needs to be chosen such that the following inclusion for the Minkowski sums

$$\Xi + \rho B_{\boldsymbol{C}} \subset H + [0, 1)^k \tag{26.2}$$

holds for some $\rho > 0$, where $B_{\boldsymbol{C}} = \{\boldsymbol{x} \in \mathbb{R}^k : \|\boldsymbol{x}\|_{\boldsymbol{C}} \leq 1\}$ denotes the unit ball of $\|\cdot\|_{\boldsymbol{C}}$. In the sequel we will first find a ρ for a given candidate set H and adjust in a second step $\underline{\zeta}$ and $\overline{\zeta}$, such that condition 26.2 is satisfied.

3 Unwrapping

For $k = 2$ the voronoi cells are intervals and we can efficiently find the cell containing a point by a hierarchic search in $O(\log(\text{number of cells}))$.

For $k = 3$ a computationally efficient way to answer this next neighbor question is to use a 2-dimensional look up table on the cost of some discretization error. In some cases there might be measurements $\boldsymbol{D}^{-1/2}\boldsymbol{U}^T\boldsymbol{P}_\perp\tilde{\boldsymbol{\varphi}}$, which are outside the correct voronoi cell. For such measurements it might be interesting to know the second best voronoi cell. Additional information like expected monotonicity of a sequence of measurements might allow to decide whether to take first or second choice or to reject both. In figure 26.2 the lookup tables of a simple example are illustrated.

For higher dimensions the storage necessary for a lookup table might be prohibitive. A next neighbor search using e.g. a Kd-tree of $\boldsymbol{P}_\perp(-H)$ seems to be a better solution.

4 Choosing wavelengths

There are two competing goals. The unwrapping of the period numbers should tolerate noise in the phase measurement and the measure-

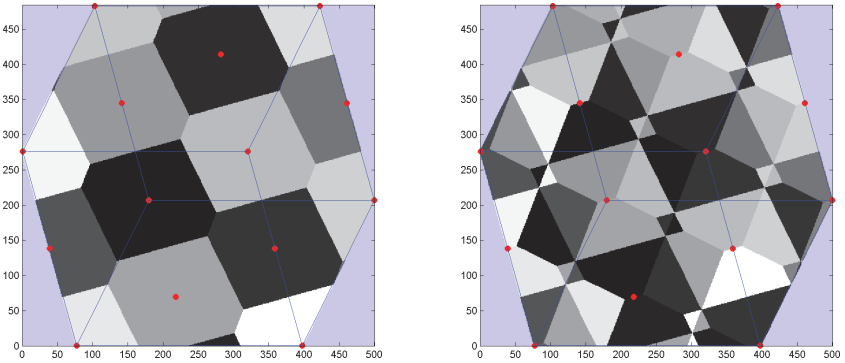


Figure 26.2: Look up tables with 500 columns ($\hat{\boldsymbol{\eta}}$ and second best $\boldsymbol{\eta}$) for the example from figure 26.1, where we assume that $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{I})$.

ment uncertainty with respect to the length of Ξ should be small, which means many periods. Let us define this more precisely.

We denote by

$$\begin{aligned} d &= \min\{\|x_i - x_j\| : x_i, x_j \in H_{\text{ref}}, x_i \neq x_j\} \\ &= \min\{\|\boldsymbol{P}_\perp \boldsymbol{\eta}\|_{\boldsymbol{C}} : \boldsymbol{\eta} \in (H - H) \setminus \{0\}\} \end{aligned}$$

the smallest distance between two centers of adjacent voronoi cells – equivalently the shortest edge of the delaunay triangulation. Remark that by symmetry it suffices to check $\boldsymbol{\eta}$ with $\boldsymbol{\tau}^T \boldsymbol{\eta} \geq 0$. From 26.1 we deduce that the Minkowski difference $H - H = \{\boldsymbol{\eta} \in \mathbb{Z}^k : |\boldsymbol{\tau}^T \boldsymbol{\eta}| \leq t_2 - t_1\}$. Summarizing, we set $\boldsymbol{\xi}' = (t_2 - t_1)\boldsymbol{\tau}$ and define $H_0(\boldsymbol{\xi}') = \{\boldsymbol{\eta} \in \mathbb{Z}^k : 0 \leq \boldsymbol{\xi}'^T \boldsymbol{\eta} \leq \|\boldsymbol{\xi}'\|^2\} \setminus \{0\}$. Then we obtain

$$d(\boldsymbol{\xi}') = \min\{\|\boldsymbol{P}_\perp \boldsymbol{\eta}\|_{\boldsymbol{C}} : \boldsymbol{\eta} \in H_0(\boldsymbol{\xi}')\}. \tag{26.3}$$

Remark that $\boldsymbol{\tau}$ in $\boldsymbol{P}_\perp$ is the normalized $\boldsymbol{\xi}'$. Then $\|\boldsymbol{P}_\perp \boldsymbol{\varepsilon}\|_{\boldsymbol{C}} < d/2$ ensures a correct unwrapping. Given a level of confidence $\alpha \in [0, 1]$ we want to find $\rho > 0$ such that

$$\alpha = \mathbb{P}\left(\|\boldsymbol{P}_\perp(\rho\boldsymbol{\varepsilon})\|_{\boldsymbol{C}} < \frac{d}{2}\right) = \mathbb{P}\left(\|\boldsymbol{P}_\perp \boldsymbol{\varepsilon}\|_{\boldsymbol{C}}^2 < \frac{d^2}{4\rho^2}\right) = P\left(\frac{k-1}{2}, \frac{d^2}{8\rho^2}\right),$$

where we use that $\|\boldsymbol{P}_\perp \boldsymbol{\varepsilon}\|_{\boldsymbol{C}}^2 \sim \chi_{k-1}^2$ with cumulative distribution function $P((k-1)/2, \cdot/2)$ and P denoting the regularized Gamma function.

Consequently, for noise $\rho\boldsymbol{\varepsilon}$ with

$$\rho = \frac{d}{2\sqrt{2P^{-1}((k-1)/2, \alpha)}}$$

more than α of the unwrapped periods are correct. On the other hand, the noise parallel to Ξ determines the resolution. As $\boldsymbol{\tau}^T \boldsymbol{P}_{\|}\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \|\boldsymbol{\tau}\|^2_{\boldsymbol{C}})$ we have $\|\boldsymbol{P}_{\|}\boldsymbol{\varepsilon}\|^2/\|\boldsymbol{\tau}\|^2_{\boldsymbol{C}} \sim \chi^2_1$. For given level of confidence α we search for the resolution quantity $\Delta > 0$ such that

$$\alpha = \mathbb{P}(\|\boldsymbol{P}_{\|}(\rho\boldsymbol{\varepsilon})\| < \Delta) = P\left(\frac{1}{2}, \frac{\Delta^2}{2\rho^2\|\boldsymbol{\tau}\|^2_{\boldsymbol{C}}}\right),$$

which gives

$$\Delta = \rho\|\boldsymbol{\tau}\|_{\boldsymbol{C}}\sqrt{2P^{-1}(1/2, \alpha)} = d\frac{\|\boldsymbol{\tau}\|_{\boldsymbol{C}}\sqrt{P^{-1}(1/2, \alpha)}}{2\sqrt{P^{-1}((k-1)/2, \alpha)}}.$$

This means that for noise $\rho\boldsymbol{\varepsilon}$ more than α^2 estimated points on Ξ are correctly unwrapped and differ from the noise free point by less than Δ. The resolution is characterized by the normalized quantity

$$\delta = \frac{\Delta}{\|\boldsymbol{\xi}(\bar{\zeta}) - \boldsymbol{\xi}(\underline{\zeta})\|}.$$

Hence, we need to find long $\boldsymbol{\xi}'$ such that the corresponding $d(\boldsymbol{\xi}')$ is large. There are two problem settings. For a given confidence level α we search a $\boldsymbol{\xi}'$ such that δ is smaller a given resolution and ρ, equivalently $d(\boldsymbol{\xi}')$, is maximal. Alternatively, the confidence level α and a lower bound for ρ i.e. $d(\boldsymbol{\xi}')$ are given and we search a $\boldsymbol{\xi}'$ such that δ is minimal, equivalently $\|\boldsymbol{\xi}'\|$ is maximal.

The function $d(\boldsymbol{\xi}')$ to be maximized consists of many local maxima and is not continuous, e.g. $x \mapsto d(x\boldsymbol{\tau})$ is piecewise constant and monotonically decreasing. Consequently, we have to enumerate the local maxima.

In the case of choosing optimal gratings for the infrared projection there are further constraints. The length of the grating is given by the optical setup of the projector. The smallest wavelength is restricted by manufacturing limits and the modulation transfer function of camera

and projector. For larger wavelengths the deviation from a sinusoidal pattern increases. That is why there is a largest acceptable wavelength, which is shorter than the length of the grating. For period numbers this means that each component of Ξ needs to span at least a minimal number of periods (greater one) and should not exceed a maximal number. This constrained can be written as a component wise inequality $\underline{\boldsymbol{\xi}'} \leq \boldsymbol{\xi}' \leq \overline{\boldsymbol{\xi}'}$.

Remark that having in one component only one period would be optimal for avoiding wrong unwrapping causing errors larger than the noise.

5 Geometric interpretation

The matrix $\boldsymbol{P}_\perp \boldsymbol{C} \boldsymbol{P}_\perp$ is positive semidefinite and $\boldsymbol{\tau}$ is the eigenvector of the eigenvalue 0. Consequently, $K(d, \boldsymbol{\tau}, \boldsymbol{C}) = \{\boldsymbol{x} \in \mathbb{R}^k : \|\boldsymbol{P}_\perp \boldsymbol{x}\|_{\boldsymbol{C}}^2 < d^2\}$ defines an open elliptical cylinder with axis $\boldsymbol{\tau}$. Then we can reformulate the definition of d in 26.3 as follows

$$d(\boldsymbol{\xi}') = \max\{x \geq 0 : K(x, \boldsymbol{\tau}, \boldsymbol{C}) \cap H_0(\boldsymbol{\xi}') = \emptyset\},$$

where $\boldsymbol{\tau} = \boldsymbol{\xi}'/\|\boldsymbol{\xi}'\|$. Using the eigenvalue decomposition $\boldsymbol{P}_\perp \boldsymbol{C} \boldsymbol{P}_\perp = \boldsymbol{U}_{\boldsymbol{\tau}} \boldsymbol{D}_{\boldsymbol{\tau}} \boldsymbol{U}_{\boldsymbol{\tau}}^T$ we can rewrite this as

$$d(\boldsymbol{\xi}') = \max\{x \geq 0 : K(x, \boldsymbol{\tau}, \boldsymbol{I}) \cap \boldsymbol{U}_{\boldsymbol{\tau}} \boldsymbol{D}_{\boldsymbol{\tau}}^{1/2} \boldsymbol{U}_{\boldsymbol{\tau}}^T H_0(\boldsymbol{\xi}') = \emptyset\}$$

with a circular cylinder and transformed grid points.

Let us consider the contact points $\mathrm{cl}(K(d(\boldsymbol{\xi}'), \boldsymbol{\xi}'/\|\boldsymbol{\xi}'\|, \boldsymbol{C})) \cap H_0(\boldsymbol{\xi}')$. When ever there are less than k contact points, they do not determine the cylinder, i.e. in a small neighborhood there is a $\boldsymbol{\xi}_1'$ and $K(d(\boldsymbol{\xi}_1'), \boldsymbol{\xi}_1'/\|\boldsymbol{\xi}_1'\|, \boldsymbol{C})$ has the same contact points but $d(\boldsymbol{\xi}_1') > d(\boldsymbol{\xi}')$. Consequently, the local maxima are represented by those cylinders, which have k contact points determining their diameter and axis orientation.

For $k = 2$ let $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ be the contact points of such a cylinder. Its axis passes through $(\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2)/2$. The point $\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2$ lies on the axis and is the first point inside this cylinder, consequently the end point of the cylinder. From a given cylinder one can construct two new ones choosing as contact points the pairs $(\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2, \boldsymbol{\eta}_2)$ and $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2)$. This

observation allows a recursive enumeration of the local maxima starting with $((0,1)^T,(1,1)^T)$.

For $k>2$ this appears to be more complicated. Let $(\boldsymbol{\eta}_1,\ldots,\boldsymbol{\eta}_k)$ be the contact points of a cylinder with radius r then for the normalized axis vector $\boldsymbol{\tau}$ it holds

$$\begin{aligned} r^2 &= \boldsymbol{\eta}_i^T(\boldsymbol{I}-\boldsymbol{\tau}\boldsymbol{\tau}^T)\boldsymbol{C}^{-1}(\boldsymbol{I}-\boldsymbol{\tau}\boldsymbol{\tau}^T)\boldsymbol{\eta}_i \\ &= \boldsymbol{\tau}^T(\|\boldsymbol{\eta}_i\|^2\boldsymbol{I}-\boldsymbol{C}^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T-\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\boldsymbol{C}^{-1}+(\boldsymbol{\tau}^T\boldsymbol{C}^{-1}\boldsymbol{\tau})\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T)\boldsymbol{\tau} \end{aligned}$$

for $i=1,\ldots,k$. Let $\boldsymbol{A}_i=\boldsymbol{\eta}_{i+1}\boldsymbol{\eta}_{i+1}^T-\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T$ then we get the following system of $k-1$ bi-quadratic equations

$$0=\boldsymbol{\tau}^T((\|\boldsymbol{\eta}_{i+1}\|^2-\|\boldsymbol{\eta}_i\|^2)\boldsymbol{I}-\boldsymbol{C}^{-1}\boldsymbol{A}_i-\boldsymbol{A}_i\boldsymbol{C}^{-1}+(\boldsymbol{\tau}^T\boldsymbol{C}^{-1}\boldsymbol{\tau})\boldsymbol{A}_i)\boldsymbol{\tau}.$$

For $\boldsymbol{C}=\boldsymbol{I}$ this simplifies to a system of $k-1$ quadratic equations

$$0=\boldsymbol{\tau}^T((\|\boldsymbol{\eta}_{i+1}\|^2-\|\boldsymbol{\eta}_i\|^2)\boldsymbol{I}-\boldsymbol{A}_i)\boldsymbol{\tau}.$$

The k-th equation is $1=\boldsymbol{\tau}^T\boldsymbol{\tau}$. Remark that for a solution $\boldsymbol{\tau}$ the vector $-\boldsymbol{\tau}$ describing the same cylinder is a solution, too. Depending on the contact points there is in general more than one cylinder, e.g. for $k=3$ there are up to four cylinders.

6 Application example

In the sequel let us assume that the components of the noise are independent and their variances are identical. Setting $\boldsymbol{C}=\boldsymbol{I}$ the value of ρ represents the acceptable standard deviation.

For the infrared fringe projector the task was to find good wavelength combinations for 30 to 39 periods on $k=3$ gratings. Enumerating all cylinders gives the results shown in figure 26.3. By symmetry we could restrict the enumeration to contact points $\{\boldsymbol{\eta}\in\mathbb{Z}^3:0\le\eta_1\le\eta_2\le\eta_3\}$.

In the sequel all numbers are based on the confidence level $\alpha=99\%$. As expected three pattern are able to tolerate more noise than two. The highest green point for two pattern in the left diagram of figure 26.3 represents the combination $\boldsymbol{\xi}'=(30,31)$ with $\rho\approx 0.0045$ and $\delta\approx 0.00027$.

For three pattern the most noise tolerant combination with a resolution $\delta\approx 0.001\le 0.001$ is $\boldsymbol{\xi}'\approx(31.011,31.985,37.003)^T$ with an acceptable standard deviation $\rho\approx 0.022$. The corresponding lookup table is shown in the left of figure 26.4.

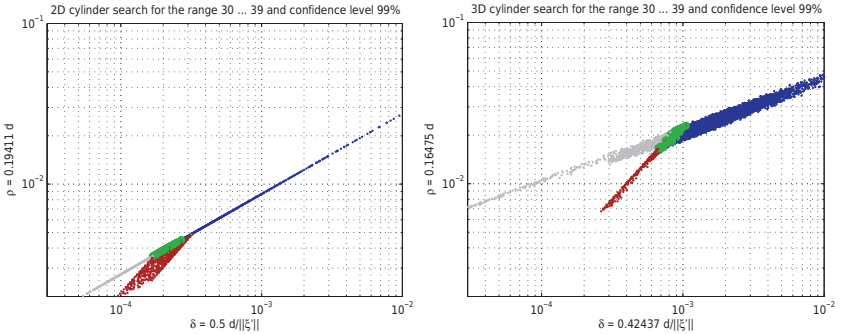


Figure 26.3: Local maxima of $d(\boldsymbol{\xi}')$ plotted in a resolution δ and standard deviation ρ diagram for a given confidence level. Blue points represent cylinders with $\xi_1' < 30$. The gray points mark cylinders with $\xi_3' > 39$. Limiting these cylinders to the range results in the red points – this corresponds in the diagram to a horizontal shift to the right. The green points stand for cylinders, which terminate inside the range.

This cylinder has the contact points $(1,1,1)^T$, $(6,6,7)^T$ and $(30,31,36)^T$. What about the influence of manufacturing tolerances. Let us consider the length of a single period. A deviation from the nominal length could be interpreted as a part of the error budget of the phase measurement. Consequently, the difference of the actual and the nominal period relative to the nominal length should be smaller than ρ.

Yet, if there is a systematic difference for all periods this would accumulate and change $\boldsymbol{\xi}'$. As a scaling does not matter we assume that ξ_1' is correct. We can change ξ_2' and ξ_3', which represents a scaling of the second and the third pattern. For an upper bound of the loss of the noise tolerance ρ we consider only the furthest contact point $(30,31,36)^T$. The angle between $(30,31,36)^T$ and $\boldsymbol{\xi}'$ corresponds to a worst case loss of 100%. Similarly, we can compute angles for 50%, 25%, etc. Intersecting the cones with these half apex angles with the ξ_1'-plane and scaling with $\mathrm{diag}(1/\xi_2', 1/\xi_3')$ gives the curves in figure 26.4. This diagram could be read as follows. Scaling the second pattern by 0.2% and the third by $0,1\%$ could reduce ρ in a worst case to 0.011.

If manufactured $\boldsymbol{\xi}'$ gets to close to $(30,31,36)^T$ one can reduce the length of the unambiguity range. There are two cylinders terminated

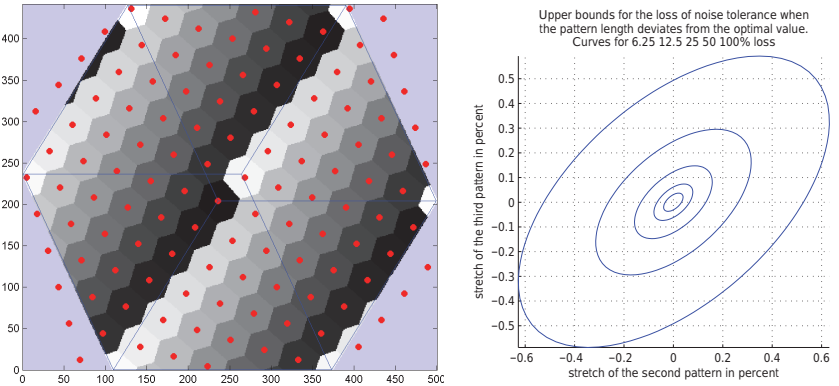


Figure 26.4: Look up table for $\boldsymbol{\xi}' = (31.011, 31.985, 37.003)^T$ with 500 columns and diagram on the impact of manufacturing deviations.

by $(30, 31, 36)^T$ with slightly better $\rho \approx 0.023$ and almost the same $\delta \approx 0.001$. They have contact points $((1,1,1)^T, (5,5,6)^T, (6,6,7)^T)$ and $((1,1,1)^T, (5,5,6)^T, (25,26,30)^T)$.

7 Outlook

The algorithm for the enumeration of the cylinders with $k = 3$ has not been optimized, yet. Interesting would be to analyze the gain in noise tolerance for $k > 3$. The algorithms for $k > 3$ are not implemented, yet.

The above cylinder search can although be interpreted differently. Given a ray from the origin with direction $\boldsymbol{\tau}$ consider the set $S_\lambda = \{\boldsymbol{\xi} \in \mathbb{Z}^k : 0 < \boldsymbol{\tau}^T\boldsymbol{\xi} \leq \lambda\}$. With increasing λ the approximation of the ray by some ray through a point in S_λ improves in the sense that $\min\{\|(\boldsymbol{I} - \boldsymbol{\tau}\boldsymbol{\tau}^T)\boldsymbol{\xi}\| : \boldsymbol{\xi} \in S_\lambda\}$ decreases. We are interested in rays, which are badly approximable for some given λ. Are there links to simultaneous approximations?

References

1. E. Lilienblum and B. Michaelis, "Optical 3d surface reconstruction by a multi-period phase shift method," *Journal of Computers*, vol. 2, no. 2, pp. 73–83, 2007.
2. T. Dunker and S. Luther, "Calibration of an infrared 3d scanner," *tm-Technisches Messen*, vol. 81, no. 1, pp. 8–15, 2014.
3. M. G. Löfdahl and H. Eriksson, "Algorithm for resolving 2 pi ambiguities in interferometric measurements by use of multiple wavelengths," *Optical Engineering*, vol. 40, no. 6, pp. 984–990, 2001.

Bildverarbeitung spielt in vielen Bereichen der Technik zur effizienten und objektiven Informationserfassung eine Schlüsselrolle. Beispielsweise in der Qualitätssicherung industrieller Produktionsprozesse und zur Fahrerassistenz haben sich Bildverarbeitungssysteme einen unverzichtbaren Platz erobert. Dennoch werden in der Bildverarbeitung weiterhin erhebliche Fortschritte gemacht: Sie werden auf der Seite der Hardware durch Weiterentwicklungen im Bereich der Sensortechnik, der Datenübertragung und durch die Zunahme der Leistungsfähigkeit von Rechnersystemen getragen. Auf der Seite der Signal- und Informationsverarbeitung sind leistungsfähige mathematische Verfahren und effiziente Algorithmen zur Verarbeitung der von Kameras erfassten Bildsignale wichtige Schwerpunkte aktueller Forschung und Entwicklung.

Der vorliegende Tagungsband des „Forums Bildverarbeitung", das am 27. und 28. November 2014 in Regensburg stattfand, greift diese hochaktuellen Entwicklungen sowohl hinsichtlich der theoretischen Grundlagen, Beschreibungsansätze und Werkzeuge als auch relevanter Anwendungen auf. Er richtet sich an Fachleute, die sich in der industriellen Entwicklung, in der Forschung oder der Lehre mit Bildverarbeitungssystemen befassen. Die Veranstaltung fand im Rahmen eines VDI/VDE-GMA-Expertenforums in Kooperation mit dem Karlsruher Institut für Technologie und dem Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung statt.

ISBN 978-3-7315-0284-5