

Szeneninterpretation unter Verwendung multimodaler Sensorik und Salienzmaßen

Zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

von der Fakultät für Elektrotechnik und Informationstechnik
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Inform. Markus Müller

aus Karlsruhe

Tag der mündlichen Prüfung:	29. Oktober 2013
Hauptreferent:	Prof. Dr.-Ing. Kristian Kroschel
Korreferent:	Prof. Dr.-Ing. Klaus D. Müller-Glaser

Erklärung

Ich versichere wahrheitsgemäß, die Dissertation bis auf die dort angegebene Hilfe selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer und eigenen Veröffentlichungen unverändert oder mit Änderungen übernommen wurde.

Markus Müller

Danksagung

Hiermit möchte ich meinen Dank an Herrn Professor Kroschel und den Korreferenten Herrn Professor Müller-Glaser ausdrücken. Unterstützung erhielt ich ferner vom Leiter des Fraunhofer IOSB, Prof. Dr.-Ing. Beyerer. Insbesondere Professor Kroschel förderte meine Arbeit in unermüdlicher Weise. Zu danken ist auch der fairen Prüfung durch die weiteren Prüfer Prof. Dr. rer. nat. Dössel (Dekan), Prof. Dr.-Ing. Puente (Prüfungsvorsitzender) und Prof. rer. nat. Lemmer. Abgesehen von den Mitarbeiterinnen und Mitarbeitern der Abteilung Videoauswertesysteme, den Gruppenleitern (Dr. Willersinn, Dr.-Ing. Meißner, Dr.-Ing. Schuchert und Herrn Heinze) sowie der Abteilungsassistentin, Frau Schreiber, gilt mein Dank Antje und meinen Kindern Clara-Luise, Wieland, Till und Hagen.

Markus Müller

Inhalt

1	EINLEITUNG.....	1
1.1	MOTIVATION	1
1.2	PROBLEMSTELLUNG UND BEGRIFFE	3
1.3	BILDHINTERGRUND.....	8
1.4	BACKTRACKING	8
1.5	KONZEPTIONELLER RAHMEN	10
2	SZENARIEN.....	12
2.1	BANKÜBERFALLSZENE	13
2.2	SZENE HOOLIGANS.....	14
2.3	SZENE FAHRZEUGSPUREN	15
2.4	SZENE HERRENLOSES GEPÄCKSTÜCK.....	15
3	SALIENZ/AUFFÄLLIGKEIT	17
3.1	SEGMENTIERUNG.....	20
3.2	SEGMENTIERUNG MIT SCHWELLWERTEN	21
3.3	STATISTISCHE HYPOTHESENTESTS	24
3.4	EXTRAKTION GLOBAL AUFFÄLLIGER STRUKTUREN	25
3.5	VARIANZEN	27
3.6	DYNAMISCHE SYSTEME.....	28
3.7	SEGMENTIERUNG MIT FUNKTIONALEN.....	29
3.7.1	<i>Detektion mittels CFAR-Ansätzen</i>	<i>31</i>
3.8	HINTERGRUNDCHARAKTERISIERUNG	32
3.9	TEXTURBASIERTE ANSÄTZE	36
3.10	NAG-AUFFÄLLIGKEITSDETEKTION	37
3.10.1	<i>Eigene Bewertung des Korn'schen Verfahrens</i>	<i>42</i>
4	EINORDNUNG UND SIGNALNAHER ANSATZ	49
4.1	PROZEDUR.....	49
5	BESCHLEUNIGUNG DER VERARBEITUNG	62
5.1.1	<i>Rechnerarchitekturen.....</i>	<i>65</i>
5.2	SIMD-REALISIERUNG DER NAG.....	66
5.2.1	<i>SIMD-Realisierung des veränderten NAG-Operators.....</i>	<i>67</i>

6	VIDEOGESTÜTZTE SICHERHEITSTECHNIK	69
6.1	VSAM – VIDEO SURVEILLANCE AND MONITORING	69
6.1.1	<i>Genereller Aufbau</i>	<i>69</i>
7	VIDEOAUSWERTESYSTEM ABUL	72
7.1	ARCHITEKTUR	73
7.2	BENUTZUNGSSCHNITTSTELLE	74
7.3	BILDAUSWERTEVERFAHREN	76
7.3.1	<i>Bildregistrierung</i>	<i>76</i>
7.3.2	<i>Bildteppiche</i>	<i>80</i>
7.3.3	<i>Bildverbesserung und Bildstabilisierung</i>	<i>84</i>
7.3.4	<i>Weitere Verfahrensansätze</i>	<i>84</i>
8	ZUSAMMENFASSUNG UND AUSBLICK	86
9	ANHANG NAG-HERLEITUNG.....	88
9.1	ZWEIDIMENSIONALE BERECHNUNG DER NAG	90
10	EIGENE ERFINDUNGEN / PATENTE.....	92
10.1	360°-PANORAMAKAMERA	92
10.2	SCANNER-TRACKER	94
10.2.1	<i>MIDI.....</i>	<i>95</i>
10.2.2	<i>Moving-Light-Geräte.....</i>	<i>95</i>
11	EIGENE PUBLIKATIONEN	99
12	LITERATURVERZEICHNIS.....	105

Symbolverzeichnis

A_i	i -tes Artefakt
\exists, \forall	Existenz- und Allquantor
\subset, \subseteq	Teilmengenbeziehungen
\emptyset	Leere Menge
\cap, \cup	Schnitt- und Vereinigungsmenge
$\ $	Betrag
x	Variable, Skalar
\mathbf{x}	Vektor
\hat{x}	Schätzwert
M	Matrix
∂	Partielles Differential
H	Homogenitätskriterium, Hypothese oder Hintergrund
\mathcal{H}^m	Hausdorff-Maß der Dimension m
Δ	Laplace-Operator: $\Delta = \nabla^2 = \sum \frac{\partial^2}{\partial x_n^2}$ oder Delta (Änderung)
∇	Nabla-Operator: $\nabla = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right)$
\mathbb{R}, \mathbb{C}	Körper der reellen/komplexen Zahlen
R_i	Bildregion i
Ω	Bildbereich

p, q	Pixel im Bildbereich
P_D	Probability of Detection
χ^2	χ^2 -Test
C	Clutter, Cluttermodell
K	Kontrastmaß
Γ	Kontur/Kante/Kantenmenge im Bildbereich
$S_I(p)$	Auffälligkeitsmaß am Punkt p einer Kontur Γ
V_T	Totale Varianz
$[\mathbf{S}, \mathbf{r}^n]$	Dynamisches System (mit \mathbf{S} Zustandsraum und \mathbf{r}^n Transformation)
Cov	Kovarianz
μ, σ	Mittelwert, Standardabweichung
m_S	Mittlere Spannweite

Abkürzungen

ABUL	Automatische Bildauswertung für unbemannte Luftfahrzeuge
Artefakt	Vom Menschen herbeigeführte Veränderung
BMBF	Bundesministerium für Bildung und Forschung
ATR	Automatic Target Recognition
CART	Camouflage Assessment in Real-time
CFAR	Constant False Alarm Rate
Clutter	Bildfiguren, die zwar zielobjektähnlich, jedoch irrelevant sind und leicht für Falschalarme sorgen
CORBA	Common Object Request Broker Architecture
COTS	Commercial Off The Shelf (kommerziell verfügbare Produkte)
CPU	Central Processing Unit
CMUPA	Carnegie Mellon University Packet Architecture
DARPA	Defense Advanced Research Projects Agency
Dbus	Desktop-Bus
DIS	Distributed Interactive Simulation (Kommunikationsprotokoll)
FhG	Fraunhofer Gesellschaft
FFT	Fast Fourier Transform
FLIR	Forward Looking Infrared (Sensor)
GIS	Geographisches Informationssystem

GNC	Graduated non Convexity (Optimierungsverfahren)
GPU	Graphics Processing Unit
IC	Integrated Circuit
IPC	Interprozesskommunikation
IR	Infrarot
KIBASX	Konfigurierbares, interaktives Bildauswertesystem, basierend auf X-Window
LWIR	Long-Wavelength IR
MIMD	Multiple Instruction, Multiple Data
MISD	Multiple Instruction, Single Data
MMO	Man-Made Object
MMX	Multi Media Extension
MWIR	Mid-Wavelength IR
NAG	Normierte Ableitung einer Gaußfunktion
OCU	Operator Control Unit
OMG	Object Management Group
OpenMP	Open Multi-Processing
ORB	Object Request Broker
POI	Point of Interest
QVGA	Quarter Video Graphics Array
QUXGA	Quad Ultra XGA
ROC	Receiver Operating Curve/Characteristic

ROE	Region of Expectation
ROI	Region of Interest
SAR	Synthetic Aperture RADAR
SCR	Signal-/Clutter-Ratio
SGA	Satellitengestützte Aufklärung
SIFT	Scale Invariant Feature Transform
SIMD	Single Instruction, Multiple Data
SISD	Single Instruction, Single Data
SPIE	Society of Photographic Instrumentation Engineers. Heute: International Society for Optical Engineering
SPU	Sensor Processing Unit
SQL	Structured Query Language
SSE	Streaming SIMD Extensions
SURF	Speed-Up Robust Features
SVGA	Super Video Graphics Array
SXGA	Super XGA
TIR	Thermisches IR (Wärmebildsensor)
UAV	Unmanned Air Vehicle
VGA	Video Graphics Array
VHS	Video Home System
VIS	Visuell-optischer Sensor (oder Visualisierung)
VSAM	Video Surveillance and Monitoring (DARPA Projekt)

WQUXGA	Wide Quad Ultra XGA
WSXGA	Wide Super XGA
WXGA	Wide XGA
XGA	Extended Graphics Array

Abbildungen

Abbildung 1: Dezember 2007: Tätlicher Übergriff auf eine Person in einer U-Bahnstation in München. Bild: dpa.	3
Abbildung 2: Beispielszene einer »Blockwelt«.....	4
Abbildung 3: IR-Beispielbild, Bild: IOSB.	6
Abbildung 4: Banküberfallszene. Bild: öffentliches Fahndungsfoto der Polizei.	7
Abbildung 5: Backtracking-Schema.	10
Abbildung 6: »A Scheme of Image Understanding«.....	10
Abbildung 7: Interpretationsszenar nach KANADE & NAGEL, mit Erweiterungen (Daniilidis, 1992).	11
Abbildung 8: Szenen von Hooligans im Fußballstadion bei rechtswidrigen Handlungen, Bild: Reuters.....	14
Abbildung 9: Leere Szene (oben links), Szene mit dynamischen Veränderungen (herumlaufende Personen, oben rechts), Frau mit einem durch die Szene gezogenen Gepäckstück/Trolley (unten links) und unbeaufsichtigtes Gepäckstück (unten rechts). Bilder: IOSB.	16
Abbildung 10: Parallel erkannte Merkmale, als Merkmalskarten dargestellt (Treisman, 1986).....	19
Abbildung 11: Demonstration des Pop-Out-Effekts bzw. einer Salienz gemäß Musterintegrationstheorie. Zur Detektion des roten Quadrats ist gemäß der FIT keine Aufmerksamkeit notwendig.	19
Abbildung 12: Beispiel für eine global auffällige Struktur eines Rings in einem Zufallsmuster.	26

Abbildung 13: Beispielbild »Detektion von Rechtecken«.....	35
Abbildung 14: Demonstration eines visuell wahrnehmbaren Texturgradienten.	36
Abbildung 15: Ideale Fahrzeugsignatur gemäß Korn'schem Ansatz links und vergrößerte Gradientenrichtungen (grauwertcodiert) rechts.	38
Abbildung 16: Senkrechtsicht auf ein Fahrzeug und korrespondierendes Gradientenrichtungshistogramm (zu beachten: das vierte signifikante Maximum liegt auf der rechten Ordinate).	40
Abbildung 17: Programmablaufplan »ROI-Screening«.....	42
Abbildung 18: Beispiel für eine »ideale« Falschdetektion (rechts). Zum Vergleich: eine ideale Fahrzeugsignatur (links).....	45
Abbildung 19: Zwei Fahrzeuge (LKW) in einem IR-Bildausschnitt. Das eine ist hell das andere dunkel. Bild: IOSB.....	46
Abbildung 20: Testbild zur Überprüfung der Kontrasteinflüsse, Orientierungen und Rauscheinflüsse.	47
Abbildung 21: Beispielhafte Anwendung des Korn-Operators (links) und des selbst entwickelten Operators (rechts) auf das Testbild.	48
Abbildung 22: Optimales Rechteckmodell: weiß auf schwarzem, homogenem Grund (links), zugehöriges Gradientenrichtungshistogramm (rechts). Das vierte Maximum liegt auf der rechten Ordinate.....	50
Abbildung 23: Vergleich der Gradientenrichtungshistogramme von lokalen Bildumgebungen und den Musterhistogrammen.	51
Abbildung 24: Vergleich – Muster und sein Histogramm (oben) sowie Zielsignatur in einem Bildbereich mit seinem Histogramm (unten).....	52
Abbildung 25: Übersicht einiger Beispielmuster mit ihren korrespondierenden Gradientenrichtungshistogrammen.	53
Abbildung 26: Muster zur Suche nach parallelen Strukturen (Straßen, Flüsse, ...) Das Histogramm weist genau zwei Maxima im Abstand von 180° auf.	54

Abbildung 27: Bild mit (aufgehellten) Erwartungsbereichen gemäß der Information »Straße«.....	55
Abbildung 28: Beispielmustervorgabe zur Suche nach Ecken im Bildbereich.....	55
Abbildung 29: Vorgabemuster für ein Flugzeug.....	56
Abbildung 30: Beispielbild (links) nach Faltung und das grauwertcodierte Gradientenrichtungsbild, rechts.....	56
Abbildung 31: Gradientenrichtungen und -beträge gleichzeitig mittels grafischer Vektoren visualisiert.	57
Abbildung 32: Anwendung »Kacheln« für Experimente bzgl. Schwellwerten von Gütemaßen, Kachelüberlappungen, etc. Kacheln mit einem überschwelligen Gütwert sind gelb gekennzeichnet. Bild: IOSB.	58
Abbildung 33: Grundwahrheit.	59
Abbildung 34: SIFT-Ansatz mit Ergebnissen und Grundwahrheit. Deutlich sichtbar ist, dass kein SIFT-Merkmal beim oberen Zielobjekt erfasst wird.	59
Abbildung 35: Ergebnis mit dem Korn-Verfahren.	60
Abbildung 36: Ergebnis mit eigenem Verfahren.....	60
Abbildung 37: Anzahl an Transistoren in CPUs (logarithmische Skala).....	62
Abbildung 38: Leistungsvergleich SISD und SIMD.....	68
Abbildung 39: Architekturübersicht des VSAM-Systems (Collins, et al., 2000).....	70
Abbildung 40: Aufbau der Kommunikationsstruktur in ABUL. Bild: IOSB.....	74
Abbildung 41: Benutzungsschnittstelle Videofenster. Bild: IOSB.	75
Abbildung 42: Benutzungsschnittstelle Auswertefenster. Bild: IOSB.....	75
Abbildung 43: Definition von Marken im Bildfluss zur Definition von relevanten Teilabschnitten. Bild: IOSB.....	76

Abbildung 44: Zwei Beispielaufnahmen derselben Szene aus unterschiedlichen Winkeln (oben), Verschiebungsvektorfeld (unten links), Registrierungsresultat in anaglyphischer Darstellung (unten rechts). Bilder: IOSB.	79
Abbildung 45: Zu beachten ist, dass bei dem Bildteppich keine Schnittkanten erkennbar sind, was an der hohen Subpixelgenauigkeit ($< 0,2$ Pixel) liegt. Bild: IOSB.	80
Abbildung 46: Bildteppiche einer Zweifarb-IR-Kamera (Auflösung: 3.000 x 900 Pixel) von Freiburg. Oben: Tagaufnahme, unten: Nachtaufnahme. Die MWIR- und LWIR-Bänder sind mit Komplementärfarben überlagert. Verwendet wurden rund 100 Einzelbilder. Einzelbilder: IRCAM GmbH; Bildteppich: IOSB.	81
Abbildung 47: Bildteppich, referenziert mit einem Ikonos-Orthobild (Hintergrund) und in blau eingezeichnetem Flugweg des Sensorträgers (gemäß GPS-Flugdaten). Bild: IOSB, Ikonos-Bild: SGA/Gelsdorf.	82
Abbildung 48: Beispiel für einen Stereobildteppich. Bild: IOSB.	82
Abbildung 49: Beispiel für einen Tiefenbildteppich. Bild: IOSB.	83
Abbildung 50: In Echtzeit generiertes Bildteppichergebnis. Bild: IOSB.	83
Abbildung 51: Bildverbesserung (u.a. Beseitigung von Pixelstörungen und Kontrastanhebung). Original: links, Korrektur: rechts. Bild: IOSB.	84
Abbildung 52: Bildgestützte Bewegtzieldetektion. Bild: IOSB.	85
Abbildung 53: Die zweidimensionale <i>NAG</i> -Funktion ($\sigma = 2$).	91
Abbildung 54: Einzelbilder aus einer Rundumerfassung der 360°-Panoramakamera..	93
Abbildung 55: Beispielhafte Veranschaulichung der Drehung von Einzelbildern.	93
Abbildung 56: Prototyp der Rundumkamera.	94
Abbildung 57: Seitenansicht eines typischen Scanners (Bild: Firma Martin).	97
Abbildung 58: Personendetektion und Tracking (Bild: IOSB).	97
Abbildung 59: Prototyp des Scanner-Tracker.	98

Zusammenfassung

Diese Arbeit rankt sich um Aspekte der Bildauswertung im Hinblick auf durch Menschen verursachte Veränderungen (Artefakte). Die Beiträge gliedern sich wie folgt:

1. Ikonisch/parametrisch¹:

Ein Vorschlag für ein Verfahren zur Auffälligkeitsdetektion (Salienz) auf ikonischer Ebene in Anlehnung an artverwandte Verfahren inkl. effizienter Implementierung durch Nutzung von Parallelisierung auf Datenebene. Das Verfahren ist unter Nutzung von SIMD-Code fertig implementiert.

2. Gesamtsystem:

Auf der Basis der Erkenntnis, dass die Nützlichkeit von automatischen Bildauswerteverfahren am zweckmäßigsten im Anwendungs- und Systemkontext demonstriert werden kann und in diesem Kontext die Entwicklung neuer Verfahren am zweckmäßigsten angestoßen wird, wurde die Entwicklung eines Bildfolgenauswertesystems für Sicherheitsaufgaben »ABUL« (automatische Bildauswertung für unbemannte Luftfahrzeuge) angestoßen.

Es wird also ein Ansatz präsentiert, der durch die sensorische Erfassung von Auffälligkeiten in exemplarischen Szenen Rückschlüsse über die sich in der Szene abspielenden oder vergangenen Vorgänge ermöglicht. Die Vorgänge zeichnen sich durch Änderungen in der Szene aus, wobei hier Veränderungen durch menschliches Einwirken gemeint sind (Erscheinen oder Verschwinden von Objekten, Personen oder hinterlassene Spuren). Bei diesen Vorgängen wird unterstellt, dass eine Nicht-Kooperativität beteiligter Akteure sowie ein dem Beobachter unbekanntes Drehbuch (Intention, Absicht) zugrunde liegt, d.h. dass zielgerichtete Aktionen stattfinden bzw. stattfanden und diese Aktionen Veränderungen in Form von Spuren hinterlassen.

¹ Bei der ikonischen Bildverarbeitung liegt das Ergebnis in bildlicher und bei der parametrischen in anderer Datenstruktur vor.

1 Einleitung

Eine Lösung von Aufgaben im Bereich der digitalen Bildverarbeitung benötigt für einen erfolgreichen Einsatz heutzutage noch eine enge Begrenzung der Aufgabenstellung.

FRAUNHOFER ALLIANZ VISION

Leitfaden zur industriellen Bildverarbeitung

1.1 Motivation

Der übergeordnete Anwendungskontext dieser Arbeit rankt sich um das Thema »Sicherheit«. Diesem Thema wird national wie international verstärkte Beachtung geschenkt. Auf nationaler Ebene äußert sich dieses z.B. durch die erstmalige Finanzierung diverser Verbundprojekte einer Programmlinie »Querschnittsorientierte Sicherheitsforschung« durch das Bundesministerium für Bildung und Forschung (BMBF) auf den folgenden Gebieten²:

- Gesellschaftliche Aspekte
- Biometrie
- Mustererkennung
- Integrierte Schutzsysteme
- Detektion von Gefahrstoffen

Ebenso gibt es vergleichbare Aktivitäten auf europäischer Ebene. Ausgehend von der »Lissabon-Strategie« der Europäischen Union soll Europa zur wettbewerbsfähigsten Region der Welt gemacht werden (BMBF, 2007). Dazu wurde das mit 54 Milliarden Euro ausgestattete und über sieben Jahre laufende 7. Rahmenprogramm

² <http://www.bmbf.de/de/18187.php>, abgerufen am 10.12.2013.

aufgesetzt, welches somit eine Budgeterhöhung im Vergleich zum Vorgängerprogramm um 60% erhielt. Im 7. Rahmenprogramm wurde die Sicherheitsforschung erstmalig auf europäischer Ebene explizit gefördert. Das in Ausarbeitung befindliche und für 2014-2020 angesetzte Folgeprogramm der EU »Horizon 2020« soll mit einem Budget in Höhe von 80 Milliarden Euro ausgestattet werden. Auch das anstehende Rahmenprogramm wird den Aspekt Sicherheitsforschung explizit adressieren (EU, 2011).

National hat sich u.a. die Fraunhofer Gesellschaft (FhG) ebenso mit großem Engagement des Themas »Sicherheit« angenommen. Das Thema wird vor allem im Kontext »Morgenstadt« gesehen, bei dem den Herausforderungen der Zukunft u.a. auch auf dem Gebiet der Sicherheit begegnet werden soll (Bullinger, et al., 2012 S. 159-180). In dem Zusammenhang ist insbesondere auch von dem Konzept »Smart Cities« die Rede, bei dem es um Städte mit einer erhöhten digitalen, vernetzten und intelligenten Informationstechnologie geht (PERSPECTIVES, 2013).

Das Fraunhofer Institut für Optronik, Systemtechnik und Bildauswertung (IOSB) hat ein eigenes Geschäftsfeld »Zivile Sicherheit« eingerichtet, in dem die Felder »Schutz des Bürgers«, »Schutz von Infrastrukturen«, »Cyber Security« sowie »Privatheit und Datenschutz« bearbeitet werden (Beyerer, et al., 2012). Auch wenn der Schutz der Persönlichkeitsrechte ein wesentliches Feld ist, um z.B. die Akzeptanz von Sicherheitstechnologien in der Gesellschaft zu erhöhen (Vagts, 2013), fokussiert sich die vorliegende Arbeit auf (system-) technische Aspekte. Beim Schutz des Bürgers geht es z.B. um frühzeitige Erfassung von im Ablauf befindlichen kritischen Situationen (z.B. gewaltsame Übergriffe), um die Chance zu wahren, noch rechtzeitig eingreifen zu können.



Abbildung 1: Dezember 2007: Tätlicher Übergriff auf eine Person in einer U-Bahnstation in München. Bild: dpa.

Neben dedizierten und möglichst flexiblen Einzelverfahren sind auch Arbeiten zur Auffälligkeitserfassung (Salienz) sowie die Realisierung von Systemkonzepten und Integrationssystemen nötig.

1.2 Problemstellung und Begriffe

Das einleitende Zitat der FRAUNHOFER ALLIANZ VISION bezieht sich auf die industrielle Bildverarbeitung. Dieses Feld zeichnet sich im Wesentlichen durch ihre Möglichkeit zur ausgeprägten Spezialisierung in dreifacher Hinsicht aus: mit spezialisierten Verfahren für eine spezialisierte Aufgabenstellung werden spezialisierte Bilder ausgewertet, d.h. es werden spezielle Sensoren ausgewählt und optimiert, der Bildhintergrund ist bekannt oder festlegbar (z.B. die Farbe vom Fließband), die Beleuchtung kann optimiert werden und soll im Betrieb unverändert (oder gezielt veränderlich) bleiben, die Welt der möglichen Gut-/Fehlsignaturen ist vergleichsweise eingeschränkt und weitere Randbedingungen, wie z.B. Taktzeiten, Lage/Orientierung und der beeinflussbare Zustand der Stückgüter sind bekannt oder festlegbar. Diese Aspekte ermöglichen die angesprochene a priori Spezialisierung der eingesetzten Verfahren unter Wegfall der Notwendigkeit einer weitgehenden Adaptivität im Betrieb

und führen dazu, dass die industrielle Bildverarbeitung in vielen Bereichen erfolgreich operieren kann:

Bildverarbeitungsverfahren sind in dem Maße erfolgreich, wie sie vor dem Betrieb spezialisiert werden können.

Prominente Beispiele einer spezialisierten Mikrowelt mit entsprechenden Aufgabenstellungen und Verfahren stellen die polyhedralen »Blockwelten« dar, die aus definierten Klötzen mit unterschiedlicher Form und Farbe bestehen, welche auf einer planen Ebene angeordnet sind (vgl. z.B. (Waltz, 1975)). Die Aufgabe für das Sichtsystem besteht darin, mit Hilfe eines Roboterarms unterschiedliche Anordnungen der Klötze zu realisieren.

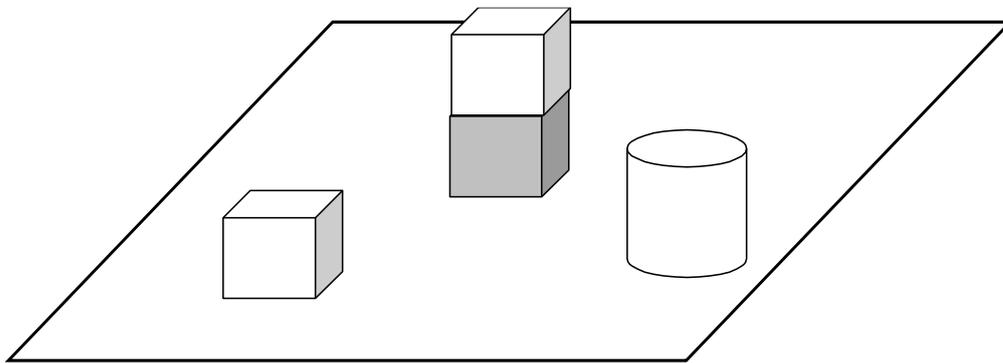


Abbildung 2: Beispielszene einer »Blockwelt«.

Demgegenüber liegen dieser Arbeit keine »spezialisierten« Bilder vor, sondern »natürliche«³, d.h. es sollen relevante Zielobjektsignaturen in ihrer angestammten, unbekanntem Umgebung gefunden werden (man könnte daher auch von »in situ«-Bildern sprechen). Die Problematik resultiert aus den größtenteils nicht bekannten bzw. unbeeinflussbaren Randbedingungen sowie der Variabilität des Dargebotenen:

- Vielfalt (Zustand, Farbe, usw.) und Anzahl unterschiedlicher Zielobjekte.
- Unbekannter Hintergrund, auf bzw. vor dem sich die Zielobjekte aufhalten.

³ Diese Unterscheidung zwischen »spezialisierten« und »natürlichen« Bildern ist (Bunke, et al., 1987) entnommen.

- Nicht kalibrierte Sensorik mit zum Teil unbekanntem oder nur grob bekannten Sensoreigenschaften.
- Ungenaue Aufnahmeparameter.
- Unbekannte oder ungenaue Wetter- und Beleuchtungsverhältnisse.
- Im Kontext von Sicherheitsaufgaben ist kein kooperatives Verhalten der Akteure zu erwarten.

Die Probleme dieser Anwendungsdomäne begründet die weitgehende Abstinenz automatischer Verfahren in aktuellen Systemen. Die Ursache für diesen Mangel besteht darin, dass die Verfahren vor und bei ihrer Anwendung nicht ausreichend spezialisiert werden können.

Teilaspekte dieser Arbeit ranken sich um die automatische Erkennung von künstlichen Objekten bzw. Artefakten (Man-Made Objects) in angestammter Umgebung, den sogenannten Primärartefakten, sowie von sonstigen Szenenstrukturen, die Hinweise liefern könnten (z.B. Spuren, Straßen), den Sekundärartefakten.

Beide zu erkennenden Strukturen werden im Folgenden zusammenfassend als Artefakt (durch den Mensch verändert) bezeichnet. Unter einem Artefakt versteht man allgemein »*das durch menschliches Können Geschaffene bzw. ein Kunsterzeugnis*« (DUDEN). Damit ist also nicht etwa das Verständnis eines Artefakts im nachrichtentechnischen Sinne als Störsignal o.ä. gemeint. Gleichwohl kann der Standpunkt vertreten werden, dass ein Artefakt stets eine Art »Störung« im Hinblick auf einen natürlichen/angestammten Hintergrund angesehen werden kann.

Die folgende Abbildung soll die verwendeten Begriffe erläutern: zu sehen ist eine Luftbildaufnahme eines IR-Sensors. In der Szene sind als Sekundärartefakte Straßen, Spuren und Stellflächen zu erkennen. Es liegen ferner zwei Primärartefakte in Form von Fahrzeugsignaturen (ein kaltes und ein warmes Fahrzeug) vor. Die Fahrzeugsignaturen, vor allem des kalten Fahrzeugs, sind kontrastschwach. Es ist ferner zu beobachten, dass die Fahrzeuge in der Szene nicht »irgendwo« stehen, sondern Hinweise auf ihren Aufenthaltsort durch die Sekundärartefakte gegeben sind.

Die Szene demonstriert, dass die Fahrzeuge als Zielobjekte auch vom Menschen nur bzw. zuverlässiger detektiert werden, wenn die durch sie generierten Veränderungen in der Szene, also die Spuren, berücksichtigt werden. Damit wird auch die Nutzung

von a priori Wissen aufgezeigt sowie generell die Einbeziehung des Kontexts bzw. der Kausalität (Drehbuch, aufeinander abgestimmte Ereignisse) verdeutlicht.



Abbildung 3: IR-Beispielbild, Bild: IOSB.

Artefakte entstehen in einer Szene bzw. im Bild durch Objekte oder aus deren *Spuren*. Eine Spur eines Objekts stellt jede Veränderung der Szene dar, die mit dem bildgebenden Sensorsystem erfasst werden kann, also eine Veränderung des (Bild-)Hintergrunds.

Mit physikalisch gegebenen Spuren im Szenenbereich kann man nur unter bestimmten Voraussetzungen rechnen (z.B. loser Untergrund). Viele Szenen spielen sich jedoch in Innenraumbereichen ab, sodass mit keiner physikalischen Spurbildung zu rechnen ist. In manchen dieser Fälle kann eine virtuelle Spur durch die Berechnung einer Trajektorie konstruiert werden. Abgesehen vom Erscheinungsbild kann auch die Analyse der (virtuellen) Spur Auffälligkeiten (z.B. im Bewegungsverhalten) aufweisen.

Die folgende Abbildung (ein Fahndungsfoto der Polizei) verdeutlicht diese Überlegungen: eine Person, die einen Banküberfall anstrebt, wird sich anders verhalten als eine Person, die ihren Bankgeschäften nachgehen möchte. Ebenso ist, wie im Beispielbild, davon auszugehen, dass die Zielperson bezüglich ihres Erscheinungsbilds

Auffälligkeiten gegenüber regulären Bankkunden aufweist (z.B. durch das Tragen einer Maske).



Abbildung 4: Banküberfallszene. Bild: öffentliches Fahndungsfoto der Polizei.

Jedes künstliche Objekt einer Szene stellt also ein *Szenenartefakt*, d.h. ein Artefakt im Szenenbereich dar. Erfasst ein bildgebender Sensor die Szene mit den enthaltenen Szenenobjekten, so resultiert ein Szenenartefakt ggf. in einem *Bildartefakt* im Bildbereich. Es liegen also folgende Korrespondenzen vor:

- a) Szenenbereich → Bildbereich (Messwerte eines bildgebenden Sensors).
- b) MMO → primäres Szenenartefakt → primäres Bildartefakt.
- c) Analog: Objektspuren als sekundäres Szenenartefakt.

Auf der Basis der Erkenntnis, dass die »Spezialisierung« ein Schlüssel zum Auswerteeerfolg ist, werden Möglichkeiten zur Spezialisierung gesucht und ausgenutzt. Damit ordnet sich diese Arbeit in den Bereich zwischen der Auswertung von hochspezialisierten Bildern (z.B. industrielle Bildauswertung) und Alltagsbildern ein.

1.3 Bildhintergrund

Neben den Zielobjekten stellt der Szenenhintergrund ein wichtiges Phänomen dar, weil Zielobjekte mit dem Hintergrund wechselwirken (z.B. durch Fahrzeugspuren). Traditionell wird dem Hintergrund nur in Sonderanwendungen eine besondere Bedeutung beigemessen (z.B. im Bereich der Tarnung und Täuschung), demgegenüber wird der Hintergrund bei herkömmlichen Bildauswertansätzen als »Störgröße« angesehen, die insbesondere aufgrund von sogenanntem »Clutter« (zielobjektähnliche Hintergrundsignatur) unerwünschte Falschdetektionen verursacht. Die meisten Ansätze konzentrieren sich auf die Modellierung der Zielsignaturen, weniger auf den Hintergrund. Diese Ansätze sind aus mehreren Gründen unzureichend: sie ignorieren die Wechselwirkungen von Zielsignatur und Hintergrund und verschenken daher wertvolle Hinweise. Die Vorbehalte gegen eine Analyse des Hintergrunds liegen an der »Closed World Assumption«, wobei argumentiert wird, die Vielfalt möglicher Zielsignaturen sei klein im Vergleich zur Vielfalt möglicher Hintergrundsignaturen, was eine Hintergrundcharakterisierung erschwere.

1.4 Backtracking

Der Begriff »Backtracking« wurde vom amerikanischen Mathematiker D.H. LEHMER (1905-1991) im Zuge seiner Studien von speziellen Primzahlen geprägt (Lehmer, 1957). Im Bereich der Informationsverarbeitung stellt Backtracking einen rekursiven Algorithmus dar, der eine Verfeinerung einer erschöpfenden *depth-first*-Suche realisiert. Als solcher findet er häufig Anwendung bei *Constraint Satisfaction*-Problemen.

Das traditionelle Vorgehen von Detektions- und Erkennungsverfahren besteht in der Bilderfassung, Vorverarbeitung, Merkmalextraktion und Klassifikation. Das entspricht einer starren, sequentiellen Methodik.

Dieses starre Vorgehen erhielt positive Erweiterungen durch *iterative* Verfahren, von denen man im Wesentlichen 3 Klassen identifizieren kann:

1. Ansätze, die diskrete Abstufungsintervalle bzgl. der Bildauflösung betrachten, mit dem prominenten (und im Bereich der Bildverarbeitung ersten) Beispiel der Auflösungspyramiden in diskreten Zweierpotenzen (Burt, et al., 1983). Allgemeiner gefasst, die diskreten Filterbänke sowie Ansätze mit kontinuierlichen Abstufungsintervallen (Skalenraumanalyse & Wavelets).
2. CFAR⁴-Ansätze, siehe Abschnitt 3.7.1.
3. Ansätze, die auf A. M. LJAPUNOV (1857-1918) zurückgehen, welcher in seiner Dissertation 1892 dynamische (iterative) Systeme studierte und welche die Markov-Modellansätze beeinflussten. Markov legte eine Grundlage für bestimmte Formen Künstlicher Neuronaler Netze (KNN), insbesondere für die rückgekoppelten Hopfield-Netze (Hopfield, 1982) und den iterativen Ansatz im Ising-Modell (Ising, 1925). Damit wurden ebenso die Grundlagen für neuere Ansätze zur Optimierung von Energie-/Entropiefunktionen gelegt (prominente aktuelle Beispiele: Mumford-Shah Energie-Funktional, Snakes und »Aktive Konturen«).

Der Begriff »Backtracking« soll sich in dieser Arbeit jedoch nicht auf lokale Iterativität ein und desselben Verfahrensansatzes beziehen, sondern im Sinne von »nochmaliger Inspektion«, ggf. mit einem anderen Ansatz (bzw. Software-Agenten). Dazu wird ein Optimierungs- bzw. Bewertungsmodul benötigt, um die Selektion und Parametrierung vornehmen zu können. Das Schema des verfolgten Backtracking-Ansatzes in der Einrichtphase verdeutlicht Abbildung 5.

⁴ CFAR: Constant False Alarm Rate.

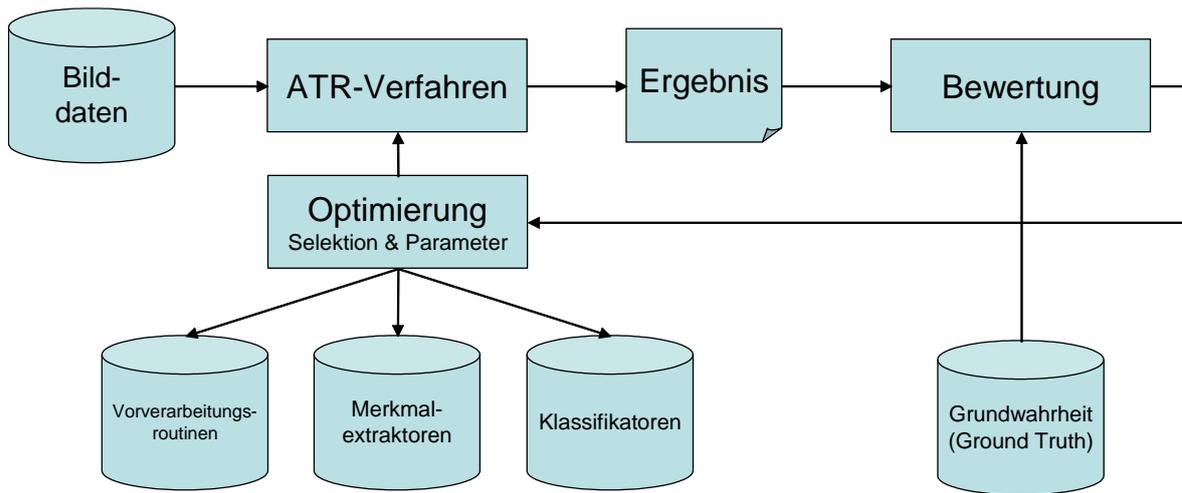


Abbildung 5: Backtracking-Schema.

1.5 Konzeptioneller Rahmen

Bereits 1978 wurde ein Systemkonzept vorgeschlagen, um aufgabenunabhängige Aspekte des Bildverstehens zu strukturieren (Kanade, 1978 S. 95-105). Dabei wird insbesondere zwischen dem Bild- und dem Szenenbereich unterschieden.

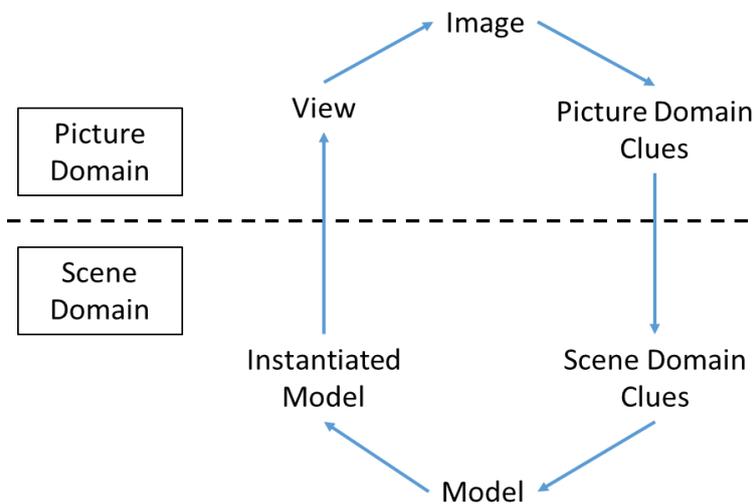


Abbildung 6: »A Scheme of Image Understanding«.

Dieser Entwurf erfuhr bald Erweiterungen (Nagel, 1979 S. 90-96) und wird auch in der Folge beachtet; z.B. (Daniilidis, 1992), (Müller, 2001).

Als »Szenenbereichshinweise« werden in dieser Arbeit Kollateral- bzw. a priori Informationen bezeichnet. »Bildbereichshinweise« liegen in Form von Ziel- und Hintergrundsignaturen vor.

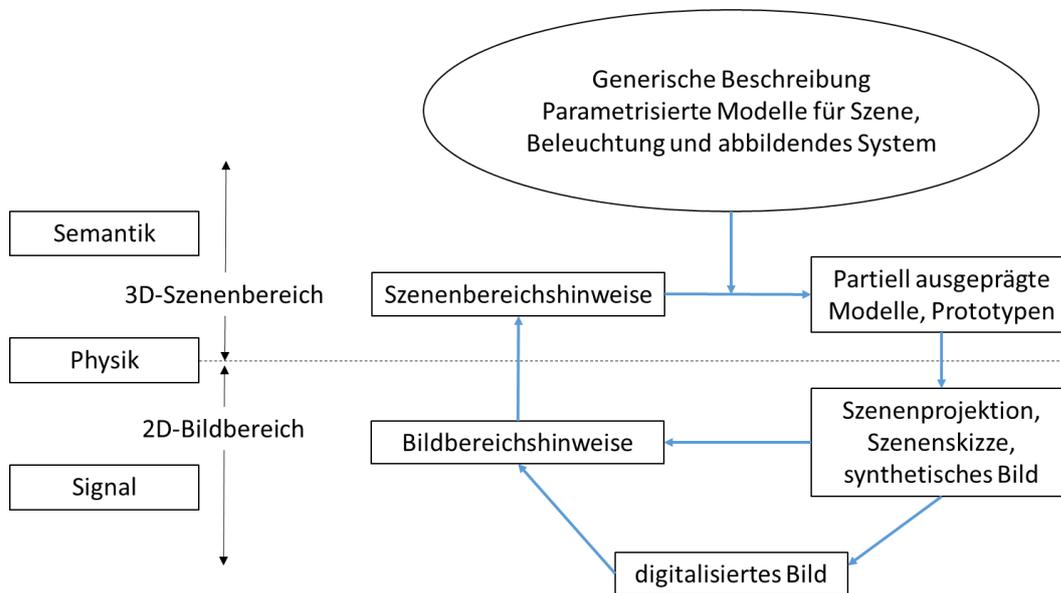


Abbildung 7: Interpretationsszenar nach KANADE & NAGEL, mit Erweiterungen (Daniilidis, 1992).

2 Szenarien

In diesem Abschnitt sollen Beispielszenarien und das auszuwählende Szenar besprochen werden. Die Zielsetzung ist, auffällige Aspekte in Szenen zu erfassen. Dabei wird davon ausgegangen, dass gewisse, sich in der Szene verändernde Bestandteile keinem Naturspiel, sondern einer *Absicht* eines Akteurs unterliegen. Problematische Absichten zeichnen sich im Gegensatz zu alltäglichen Handlungen zumeist durch *Besonderheiten* und damit eine gewisse *Auffälligkeit* bzw. eine wahrnehmungsphysiologische *Salienz* aus. Die wesentliche Charakteristik einer Salienz ist es, dass sich messbare Eigenschaften (z.B. einer Person, eines Objekts, eines Verhaltens) aus der Szene bzw. dem Szenenablauf herausheben und daher für den Menschen aber auch die Maschine leichter bzw. überhaupt erfassbar sind. Das nennt man gemäß der Musterintegrationstheorie auch »Pop-Out-Effekt«.

Die Erfassung von Auffälligkeiten stellt eine Ergänzung zur *auftragsgesteuerten* Bildauswertung dar (Monari, 2011). Bei der auftragsgesteuerten Bildauswertung wird dem Sicherheitssystem die Motivation der zu überwachenden Person direkt mitgeteilt. Z.B., wenn eine Zielperson ein Foyer einer Unternehmung betritt, sich anmeldet und einen Besuchswunsch äußert, was dann auch den weiteren, zu erwartenden Bewegungspfad der Zielperson festlegt und dem Überwachungssystem in Form eines »Auftrags« a priori mitgeteilt wird.

Klassische Bildauswertansätze versuchen, die Zielobjekte direkt zu erfassen und sehen Begleitstrukturen, wie z.B. Spuren, zurückgelassene bzw. weggeworfene Dinge/Gegenstände, etc. als irrelevante, ggf. störende Artefakte an, die es herauszufiltern gilt. Vorliegend sollen jedoch genau diese Artefakte als Hinweislieferant auf die Szenenbestandteile dienen, die als relevant anzusehen sind.

Spuren können physikalischer Natur sein, wie z.B. Fuß- und Fingerabdrücke oder die Spuren von Reifen auf entsprechendem Untergrund. Um physikalisch hinterlassene Spuren zu detektieren, werden spezialisierte (Software-) Agenten benötigt, welche die jeweils dedizierte Bildauswertaufgabe übernehmen. Dabei sind folgende Einflussfaktoren zu berücksichtigen:

- Art der Spur (Reifenspuren, Finger- oder Fußabdrücke)
- Sensorcharakteristik (insbesondere spektraler Erfassungsbereich und optische Parameter)
- Aufnahmeparameter (u.a. Aufnahmewinkel, Entfernung...)
- Szenencharakteristik (u.a. Beleuchtung, Hintergrund, Geschehen...)

Es wird mindestens ein spezialisierter Ansatz (»Softwareagent«) zur Erfassung einer speziellen Auffälligkeit benötigt. Aus der offensichtlichen Variabilität der Einflussfaktoren folgt, dass ein Erfassungssystem, welches viele unterschiedliche Situationen verarbeiten können soll, rasch äußerst umfangreich wird. Es ist also, ähnlich wie bei der auftragsorientierten Szenenerfassung, erforderlich, einen Systemansatz inklusive Systemarchitektur zu realisieren, der eine möglichst hohe Generalisierbarkeit und Erweiterbarkeit aufweist. Als Systemansatz wird daher vorgeschlagen:

1. Implementierung verschiedener Auffälligkeitsdetektoren, die generelle »Auffälligkeiten« im Bild erkennen sollen.
2. Ist eine Auffälligkeit detektiert worden, werden in Abhängigkeit der Aufgabenstellung entsprechende spezialisierte Ansätze parametrisiert und gestartet. Die Parametrisierung beinhaltet insbesondere den Ort in der Szene, in dem sich die Auffälligkeit abspielt.
3. Die spezialisierten Softwareagenten liefern Hinweise inklusive Wahrscheinlichkeiten und Gütemaße.
4. In Abhängigkeit von der getroffenen Auswahl könnten wiederum spezialisierte Softwareagenten parametrisiert und gestartet werden, die sich um eine weitergehende Analyse kümmern.

2.1 Banküberfallszene

Eine Person, die einen Banküberfall anstrebt, wird sich anders verhalten und bewegen als eine Person, die ihren gängigen Bankgeschäften nachgehen möchte. Ebenso ist zumeist davon auszugehen, dass auch schon das Erscheinungsbild der Zielperson Auffälligkeiten (z.B. eine Maske) aufweist.

Realweltliche Banküberfallszenen basieren häufig auf schlechter Videobildqualität (oftmals immer noch nur Grauwertkameras) und an die eigentlichen Bildfolgen ist sehr schwer heranzukommen. Wenn, dann kann man nur mit sehr kurzen Clips rechnen. Zumeist liegen jedoch nur Einzelbilder in der Form von Fahndungsfotos vor.

Die dem Fraunhofer IOSB vorliegenden Bild- und Videodaten vom BKA (Bundeskriminalamt), siehe Abbildung 4, unterliegen der Geheimhaltung und dürfen nicht veröffentlicht werden. Daher scheiden sie für weitere Betrachtungen aus.

2.2 Szene Hooligans

Hooligans sind gewaltbereite Personen, die sich (z.B. im Fußballstadion) im Vergleich zu Fußballinteressierten auffällig bewegen bzw. auffällige Handlungen vollziehen. So kann ein Hooligan sich aus der Menge seiner Kumpane herausbewegen, eine Handlung vornehmen (z.B. einen Gegenstand werfen) und sich wieder in die Menge seiner Kumpane zurückziehen. Auch die Erfassung solcher Szenen ist von hoher Bedeutung für Sicherheitskräfte (Stadionpersonal, Polizei), da zahlreiche Auffälligkeiten in ihnen zu finden sind. Das ist erkennbar an den globalen Auffälligkeiten bzgl. (hochfrequentem) Bewegungsverhalten der Hooligans sowie z.B. der Entzündung von bengalischen Feuern (Abbildung 8). Dem Fraunhofer IOSB liegen auch für diese Szenen Beispielvideos vor, die jedoch für eine Veröffentlichung (insbesondere aus datenschutzrechtlichen Gründen) nicht freigegeben sind.



Abbildung 8: Szenen von Hooligans im Fußballstadion bei rechtswidrigen Handlungen, Bild: Reuters.

2.3 Szene Fahrzeugspuren

In Abbildung 3 sind zwei sehr kontrastschwache Zielobjekte präsent, deren Signatur man auch als Mensch kaum entdecken kann. Von den eigentlichen Zielobjekten geht also keine Salienz aus. Auffällig sind jedoch die Spuren im Gelände. Der Mensch, so er mit der Suche nach den Fahrzeugen beauftragt ist, würde entlang der sichtbaren Spuren suchen, denn nur dort können sich die Fahrzeuge aufhalten. Die Spuren stellen also keine Störeinflüsse dar, sondern sie sind vital, um die Entdeckungsaufgabe zu lösen.

Entsprechende Daten liegen dem IOSB vor und werden beispielhaft verwendet. Sie sind auch deshalb geeignet, weil damit die Integration in das ABUL-System, welches später noch vorgestellt wird, naheliegender ist.

2.4 Szene herrenloses Gepäckstück

Gegeben ist eine Szene mit Publikumsverkehr. Es handelt sich also um eine dynamische, sich ständig verändernde Szene. Dabei sind die meisten Veränderungen irrelevant. Ziel hier ist es, eine *statische* Veränderung in einer dynamischen Szene zu detektieren. Vorliegend wird ein Gepäckstück (Trolley) abgestellt und dann verlassen. Bevor dieses Gepäckstück abgestellt wird, muss es von einer Person zunächst an den Abstellort verbracht werden. Da sich diese Szene nicht in natürlicher Umgebung abspielt, werden keine Spuren von der Person hinterlassen. Dennoch können Spuren in Form von Trajektorien, generiert durch Multi-Target-Tracker, computerintern simuliert werden.

Diese Detektion herrenloser Gepäckstücke in dynamischen Szenen entspricht der Detektion eines Diebstahldelikts, d.h. der Entfernung eines Objekts aus einer (dynamischen) Szene.

Die Szenen wurden von freiwilligen Mitarbeiter(innen) des Fraunhofer IOSB gestellt und mit entsprechender, eigener Sensorik aufgenommen.



Abbildung 9: Leere Szene (oben links), Szene mit dynamischen Veränderungen (herumlaufende Personen, oben rechts), Frau mit einem durch die Szene gezogenen Gepäckstück/Trolley (unten links) und unbeaufsichtigtes Gepäckstück (unten rechts). Bilder: IOSB.

3 Salienz/Auffälligkeit



ANZEIGE VON FRED R. BARNARD
IN »PRINTER'S INK«, 1927

Zunächst wird auf den Begriff der »Auffälligkeit« eingegangen. Der später vorzustellende Ansatz beruht in Teilen auf Vorarbeiten von (Korn, 1988), die u.a. in Bezug auf die Auffälligkeitsdetektion in (Korn, 1995) konkretisiert und u.a. in (Korn, et al., 1997) bzw. (Müller, et al., 1999) auf Anwendungen bezogen wurden. Neben den Korn'schen Ansätzen werden im Folgenden weitere Ansätze erläutert und einer einordnenden Kritik unterzogen.

Grundsätzlich sind die Begriffe »Salienz« bzw. »Auffälligkeit« nicht anwendungsunabhängig definierbar. Auffälligkeiten von Objekten in Einzelbildern zeichnen sich durch eine Kontrastierung gegen den Hintergrund aus (z.B. aufgrund ihrer Form, Farbe...), wohingegen sich Auffälligkeiten in Bildfolgen häufig durch Änderungen in der Szene über die Zeit (z.B. ein sich bewegendes Objekt) auszeichnen. Komplexer wird der Fall bei mobilen Sensorträgern, da dann ständig Änderungen im Bildbereich auftreten, die aber nicht mit einer Änderung im Szenenbereich korrespondieren. In diesen Fällen ist die Eigenbewegung des Sensors zu schätzen und entsprechend herauszurechnen. Schließlich gibt es auch noch den Fall von Änderungen in Szenen, wobei jedoch nur an bestimmten, gemäß eines Kriteriums als auffällig angesehenen Änderungen (z.B. statischen) ein Interesse bestehen könnte. Als Beispiel für eine statische Änderung in einer dynamischen Szene seien das Abstellen und der Verbleib eines Fahrzeugs im Szenenbereich genannt.

Eine Änderung/Auffälligkeit kann sich auf folgende Aspekte beziehen:

- a) Episodische Änderungen, d.h. ein kurzer Abschnitt in einem längeren Prozess.
- b) Abrupte bzw. spontane Änderungen.
- c) Repetitiv, d.h. wiederkehrende, sich wiederholende Änderungen.
- d) Phänologisch⁵, d.h. Änderungen, die sich im natürlichen Jahresablauf ergeben.

Der Pop-Out-Effekt wurde schon angesprochen. Bei diesem Effekt springt ein Zielobjekt (»Target«) aus mehreren Ablenkbildern (Distraktoren, Clutter) heraus. Das aufgrund seiner Zitierhäufigkeit wahrscheinlich bekannteste Modell »FIT – Feature Integration Theory« geht von einem mehrstufigen Wahrnehmungsprozess aus, was gleichermaßen von Psychologen, Neurophysiologen und Informatikern anerkannt wird (Treisman, et al., 1980). Die Umsetzung entsprechender Computermodule finden sich z.B. in (Itti, 2003).

Die FIT geht davon aus, dass der Mensch in der frühen Phase der Merkmaldetektion vier Merkmale wahrnimmt (Farbe, Orientierung, Größe und Distanz), ohne seine Aufmerksamkeit zu steuern. Die Aufmerksamkeit wird dann dazu verwendet, Merkmale zu koppeln. Dabei wird Objekt für Objekt einzeln auf der Basis seines Ortes codiert. Gekoppelt werden alle Merkmale, die an dem jeweiligen Zielort erkannt werden. Die Merkmale werden durch spezialisierte Module parallel codiert. Die Module bilden zwei Arten von Karten:

- a) Merkmalskarten (engl. *Feature Maps*): In diesen Karten wird festgehalten, ob die Präsenz eines Merkmals irgendwo in der Szene (bzw. im Bild) gegeben ist. Räumliche Aspekte werden nur implizit festgehalten. Eine solche Karte gibt keine Aussage über die Lokalisierung oder darüber, über welche anderen Merkmale ein z.B. rotes Objekt noch verfügt.
- b) Ortskarte (Hauptkarte): Diese Karte codiert, wo Merkmale lokalisiert sind, jedoch nicht welche Merkmale wo zu finden sind.

⁵ »Phänologie«: Lehre von den Erscheinungen des jahreszeitlichen Ablaufs in der Pflanzen- u. Tierwelt (DUDEN, 21. Aufl.)

Man benötigt also einen Weg, um Merkmale zu lokalisieren und zusammengehörige Merkmale zu identifizieren. Die fokale Aufmerksamkeit soll diese Aufgabe übernehmen. Die Aufmerksamkeit wandert in der Ortskarte umher (via Fokus) und wählt die Merkmale aus, die mit dem jeweiligen Ort des Betrachtungsfokus assoziiert sind. Merkmale anderer Objekte werden dabei ausgeklammert. Wahrgenommene Merkmale bilden dann eine temporäre Objektrepräsentation.

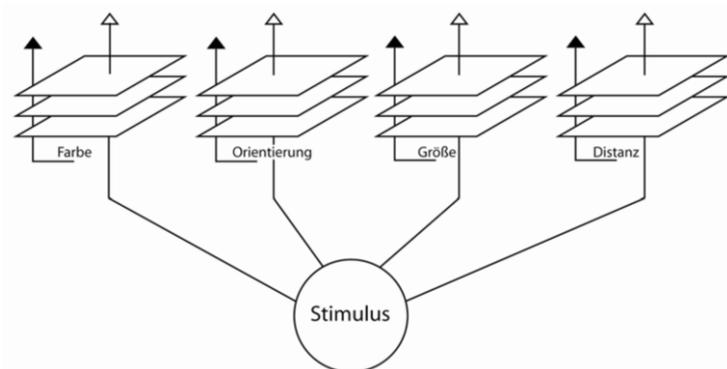


Abbildung 10: Parallel erkannte Merkmale, als Merkmalskarten dargestellt (Treisman, 1986).

In weiteren Phasen werden (ggf. weitere) erkannte Merkmale kombiniert und damit eine Erkennung durchgeführt.

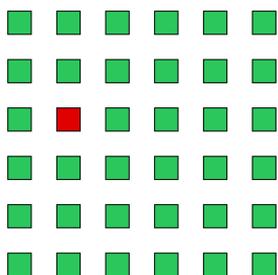


Abbildung 11: Demonstration des Pop-Out-Effekts bzw. einer Salienz gemäß Musterintegrationstheorie. Zur Detektion des roten Quadrats ist gemäß der FIT keine Aufmerksamkeit notwendig.

Wenn von »Auffälligkeitsdetektion« die Rede ist, dann ist damit die Kennzeichnung von einzelnen Bildregionen als »auffällig« gemeint. Es handelt sich also im Grunde

um einen Segmentierungsschritt in einem frühen Stadium eines Bildinterpretationsprozesses. Das Bild soll in »auffällige« (d.h. für die jeweilige Aufgabenstellung wahrscheinlich relevante) und »unauffällige« Regionen unterteilt werden. Die als »auffällig« angesehenen Bildbereiche werden als ROI (Region of Interest) bezeichnet und nur sie dienen als Vorlage für eine weitere, gründlichere Inspektion durch maschinelle oder menschliche Partizipanten am Auswertungsprozess. Jeder Segmentierungsansatz benötigt ein Homogenitätsprädikat P , um eventuelle Einteilungen vornehmen zu können.

3.1 Segmentierung

Eine Segmentierung soll das Bild in sinnvolle Regionen unterteilen. Beispielsweise in die Bereiche »Primärartefakt« (Ziel), »Sekundärartefakt« (Spur) und »Hintergrund«. Grundlegend ist dabei die Annahme, dass jedes Bildsegment einem vorgegebenen Homogenitätskriterium H genügt. Die anwendungsbezogenen Ziele ranken sich um: a) Zielsignaturen vom Bildhintergrund zu trennen, b) sich überlappende oder angrenzende Zielsignaturen zu trennen oder c) eine (größere) Zielsignatur in wesentliche Bestandteile zu zerlegen. Vorliegend ist a) relevant. Mathematisch definiert sich die vollständige Segmentierung des Bildbereichs $\Omega \subset \mathbb{R}^2$ in disjunkte Regionen $R_i \subset \Omega$, für $i = 1, \dots, n$ durch:

$$\Omega = \bigcup_{i=1}^n R_i, \quad R_i \cap R_j = \emptyset, \forall i \neq j \quad (1)$$

Dabei gilt:

1. R_i zusammenhängend $\forall i$
2. $H(R_i) = true \forall i$
3. $H(R_i \cup R_j) = false \forall i \neq j$

In das Homogenitätskriterium H fließt auch schon bei trivialen Ansätzen (z.B. bei einer bimodalen Verteilungsannahme) zumindest implizites Vorwissen ein.

Auffällig ist, was von einer Art »Norm« abweicht. Gehen wir von Objekten in natürlicher Umgebung aus, dann stellt die natürliche Umgebung die »Norm« dar und

die interessierende Bildfigur die »Auffälligkeit«. Analog wie beim Begriff der »Ähnlichkeit« lässt sich auch der Begriff der Auffälligkeit nicht absolut, sondern nur aufgabenbezogen definieren. (Korn, 1995) schlug zur Definition der Auffälligkeit vor:

Wenn sich ein Objekt aufgrund seiner Bewegung, seines Kontrastes, seiner Farbe, seiner Form oder seiner 3D-Abmessungen signifikant von seiner Umgebung unterscheidet, bezeichnen wir das Objekt als auffällig.

Mathematisch sehen wir die Detektion auffälliger Bildbereiche als eine spezielle Form der Bildsegmentierung an⁶, d.h. die Unterteilung des Bildbereichs $\Omega \subset \mathbb{R}^2$ in die disjunkte Region »Hintergrund« H sowie die disjunkten Regionen »Auffälligkeit« bzw. Artefakt A_i , mit $\{H, A_i \subset \Omega\}_{i=1,\dots,m}$, für die gilt:

$$\Omega = H \cup \bigcup_{i=1}^n A_i, \quad H \cap A_i = \emptyset \quad \forall i \neq j \quad (2)$$

Die als »auffällig« angesehenen Bildbereiche werden als ROI bezeichnet und sie dienen als Vorlage für weitere, gründlichere Inspektionen durch maschinelle (oder menschliche) Auswertungen. Die Durchmusterung der Bilddaten im Hinblick auf eine Auffälligkeitsdetektion wird *Screening* genannt. Das noch vorzustellende Detektionsverfahren beruht in Teilen auf Vorarbeiten von (Korn, 1988), die in Bezug auf die Auffälligkeitsdetektion in (Korn, 1995) konkretisiert und u.a. in (Müller, et al., 1999) auf Anwendungen bezogen wurden.

3.2 Segmentierung mit Schwellwerten

Schwellwertverfahren stellen die einfachste Variante zur Segmentierung dar. Ein Bildpunkt (bzw. eine Bildregion) gilt als »auffällig«, wenn er gemäß eines vorgegebenen oder automatisch ermittelten Wertes c überschwellig ist. Das Resultat bei der Anwendung von einem Schwellwert ist ein Binär- bzw. Zweipegelebild. Sind mehrere

⁶ Es gibt viele andere Vorstellungen bzgl. der Auffälligkeitsdetektion, z.B. reicht es in trivialen Fällen, wenn lediglich die Zielkoordinate angegeben wird.

ortsunabhängige Schwellwerte ermittelt worden, so spricht man von einem Äquidensitenbild. Schwellwertverfahren können wie folgt eingeteilt werden:

- Globaler (ortsunabhängiger) oder lokaler Schwellwert
- Ein- oder mehrdimensional ermittelter Schwellwert
- Harte oder Semischwellwertbildung
- Ein oder mehrere Schwellwerte (Äquidensiten)

Kontrastschwellwert

Ein Schwellwert bzgl. Kontrast (Modulation) setzt voraus, dass bekannt ist, dass alle Zielobjekte entweder heller oder dunkler als der Hintergrund sind. Der Kontrast bzw. die Modulation der Bildfunktion f nach (Hecht, 2009) ist in (3) angegeben. Errechnet wird ein Maß dafür, wie deutlich lokale Schwankungen von der Grundhelligkeit zu unterscheiden sind.

$$K = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad (3)$$

In der Praxis nähert man diesen Wert an indem man die minimale und maximale gemessene Intensität (Grauwert) einsetzt. Diese Kontrastapproximation besitzt jedoch den Nachteil, dass schon einzelne Ausreißer (Übersättigungen, Punktrauschen, Sensor-/Abtastfehler, usw.) den approximierten Kontrastwert stark verfälschen können. Ein (maximaler) Kontrast von 1 ergibt sich bei schwarzem Hintergrund, wenn nur ein einziger Bildpunkt den maximal möglichen Grauwert besitzt.

Kontrastschwellen mit Ausreißertest

Es ist naheliegend, dass ausreißerbehaftete Messdaten eine erhebliche Abweichung bzgl. ihrer Kenndaten (Kontrast, Mittelwert, Standardabweichung, usw.) aufweisen können. Auch alle aus den Kenndaten abgeleiteten Testgrößen sind dann verfälscht. Als Ausreißer entdeckte Messwerte müssen daher aus dem Datenmaterial entfernt werden. Vom stochastischen Standpunkt aus gesehen stellen Ausreißer Abweichungen aufgrund systematischer und nicht zufallsbedingter Einflüsse dar.

Der Ausreißertest nach GRAF & HENNING berechnet das arithmetische Mittel μ und die Standardabweichung σ aus den Daten ohne die ausreißerverdächtigen Werte. Ein

Wert gilt als Ausreißer, wenn er außerhalb des Bereichs $\mu \pm 4\sigma$ liegt. Zur Automatisierung betrachtet man iterativ die größten und kleinsten Messwerte. Sobald keiner von beiden mehr entfernt wird, gilt das Datenmaterial als ausreißerfrei.

Kontrastschwellen mit Grauwertübergangsmatrizen

Bessere Ansätze beziehen die Ortsfrequenz zur Kontrastapproximation mit ein, da sich die Anordnung von hellen und dunklen Bereichen in der Ortsfrequenz niederschlägt. **Grauwertübergangsmatrizen** (engl. *co-occurrence matrix*) W_R berechnen die Häufigkeit des Auftretens von zwei Grauwerten bzgl. einer Relation (meistens eine Nachbarschaftsrelation):

$$W_R = g_1, g_2 = a_{g_1, g_2}, \quad (4)$$

mit R Relation, g_i Grauwerte und a Häufigkeit des Auftretens der Grauwerte bzgl. R . In (Haralik, et al., 1973) wurde in diesem Zusammenhang (5) vorgeschlagen.

$$K = \sum_{i=0}^{g_{\max}-1} \sum_{j=0}^{g_{\max}-1} \frac{a_{i,j}}{M 1 + |i-j|^2}, \quad (5)$$

mit g_{\max} maximal möglicher Grauwert (z.B. 255), M Anzahl möglicher Vergleiche. Je weiter $a_{i,j}$ von der Hauptdiagonalen entfernt ist, desto geringer ist sein Gewicht für den Kontrast K . Diese örtliche Abhängigkeit des Kontrastes und Richtungsabhängigkeiten können durch die Wahl der Relation R berücksichtigt werden.

Histogrammodalität

Diese Ansätze setzen voraus, dass das Grauwert histogramm eine gut ausgeprägte Modalität aufweist. Das bedeutet, dass die Grauwerte des Hintergrundbereichs nicht (oder kaum) im Objektbereich auftreten (und vice versa). Liegt ein bimodales Histogramm vor (Zielobjekt z.B. generell heller als Hintergrund), so stellt der Schwellwert c das Minimum zwischen den beiden Maxima dar. Bei multimodalen Histogrammen (Zielobjekt besteht aus Teilen, die jeweils heller als Hintergrund sind) wird analog nach den $n - 1$ Minima zwischen den n Maxima gesucht und es werden entsprechende Äquidensiten gebildet. Die Äquidensiten werden auch zur Textursegmentierung verwendet. Der Schwellwert ergibt sich aus Texturmaßen.

Dynamische Schwellwertgenerierung

Diese Verfahren lassen zwar zu, dass die Intensitäten des Hintergrundes inhomogen sind und auch die Intensitäten des Zielobjekts schwanken, jedoch wird von einer sich eher langsamen Änderung der Grauwerte ausgegangen und zusätzlich, dass sich in jedem kleineren Bereich bimodale Histogramme ausprägen.

Ortsabhängige Verfahren (Texturmaße)

Bei den meisten Schwellwertverfahren wird ein bildpunktbezogenes Charakteristikum gewählt (meist: Intensität). Jedoch können auch Oberflächenstrukturen (Textur) bzw. entsprechend angepasste Homogenitätskriterien (Texturmaße) als Segmentierungsmerkmal verwendet werden. Häufige Texturmaße sind:

Streuung als Maß für die Rauigkeit einer Textur.

Gradienten: Berechnung der Mittelwerte der Beträge und der Richtungen. Die »Gerichtetheit« einer Textur kann durch die Auswertung der Häufigkeitsverteilung der Gradientenrichtungen ermittelt werden. Zur Bestimmung einer Vorzugsrichtung reicht oft schon der Mittelwert der Gradientenrichtungen.

Kontrast (Modulation): Die Probleme mit der Kontrastapproximation (insbesondere im Hinblick auf Ausreißer) sind bereits angesprochen worden.

Die Schwellwertverfahren zur Erzeugung von Binärbildern gehen von der Prämisse aus, dass das zu suchende Zielobjekt generell »heller« (oder) »dunkler« als der Hintergrund ist. Mittels Bildinversion können beide Fälle abgedeckt werden.

3.3 Statistische Hypothesentests

Bei dem Ansatz mittels statistischem Hypothesentest geht man davon aus, dass natürliche Hintergründe (bzw. deren Textur) im Gegensatz zu Artefakten eher »zufälligen« Charakter haben. Während man bei einer Folge von Zahlen relativ leicht nachweisen kann, dass die Folge nicht zufällig ist, fällt der Nachweis, dass eine Folge zufällig ist, schwerer.

Für den Test, ob eine rechnergenerierte Zahlenfolge nicht zufällig ist, wird sehr häufig der χ^2 -Test verwendet (Sedgewick, et al., 2011). Dieser Test prüft, ob eine zufällige Verteilung der Zahlen vorliegt oder nicht (6). Das Bild ist in überlappende Parzellen aufzuteilen, wobei eine Parzelle die Größe des gesuchten, beliebig orientierten Artefakts hat. Für jede Parzelle wird der Test angewendet.

$$\chi^2 = \frac{\sum_{0 \leq i < r} h_i - \frac{N}{r}}{\frac{N}{r}}, \quad (6)$$

mit der Anzahl N aller Bildpunkte in einer Parzelle und $N < r$ (maximal vorkommender Pixelwert plus 1) sowie h_i der Häufigkeit des Pixelwerts i . Falls die χ^2 -Statistik außerhalb des Bereichs von $2\sqrt{r}$ liegt, wird der Parzelle die Hypothese »Artefakt enthalten« unterstellt.

3.4 Extraktion global auffälliger Strukturen

Als »global auffällige Strukturen« bezeichnet (Ullman, 1995) solche Bildfiguren, welche die Aufmerksamkeit des Betrachters unmittelbar auf sich lenken. Und zwar, ohne dass das Bild vom Beobachter systematisch untersucht werden müsste. Ullman unterscheidet lokal auffällige Strukturen von global auffälligen Strukturen. Erstere zeichnen sich durch auffällige Unterschiede gegenüber benachbarten Elementen aus (z.B. Farbe, Orientierung, Kontrast oder Krümmungsverhalten). Letztere bestehen aus Einzelelementen, die jeweils für sich genommen keine lokale Auffälligkeit besitzen, deren Arrangement jedoch eine »figurähnliche« und somit auffällige Gestalt annehmen. Abbildung 12 zeigt dazu beispielhaft ein Arrangement lokal unauffälliger Einzelelemente, die in einer global wahrnehmbaren Auffälligkeit resultieren. Die Abbildung lehnt sich an einen Vorschlag von (Ullman, 1995) an, kommt jedoch mit weniger Strichelementen aus.

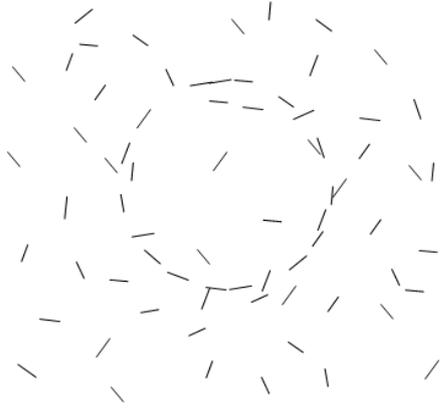


Abbildung 12: Beispiel für eine global auffällige Struktur eines Rings in einem Zufallsmuster.

Im Folgenden geht Ullman von Konturbildern aus, die z.B. aus extrahierten Linien oder Kanten zugrundeliegender Bilder gewonnen wurden. Das Verfahren definiert zunächst ein Auffälligkeitsmaß, welches sich mit der Konturlänge vergrößert und welches sich bei Krümmung der Kontur verkleinert. Darüber hinaus wird ein Schema vorgeschlagen, welches Kantenelemente zu erweiterten Konturen zusammenfügt. Die Gruppierung von Konturelementen zu (glatten) Kurven ähnelt dem Begriff der »guten Fortsetzung« der Gestaltschule im Kontext der Wahrnehmungspsychologie.

Sei Γ eine einzelne (ggf. unterbrochene) Kontur beliebiger Länge im Bild. Diese Kontur bestehe aus einer Kette kleinerer Konturelemente. Sei p ein Punkt der Kontur, dann empfiehlt sich (7) als das Auffälligkeitsmaß am Punkt p :

$$S_{\Gamma}(p) = \sum_i w_i \varphi_i, \quad (7)$$

wobei φ_i die lokale Auffälligkeit des i -ten Elements entlang der Kurve darstellt. Im einfachsten Fall könnte φ_i den Wert 1 annehmen, wenn das Konturelement präsent ist und ansonsten (d.h. bei einer Lücke) den Wert 0. Komplexere Berechnungen von φ_i könnten z.B. die Länge des betrachteten Konturelements berücksichtigen. Die Gewichte w_i des i -ten Elements berechnen sich aus $w_i = e^{-c_i}$, mit $c_i = \int K^2$ und K als Krümmung an dem betrachteten Punkt.

Die globale Auffälligkeit summiert über die Länge der Kontur und ihrem Krümmungsverhalten diese lokalen Werte auf. D.h. c_i stellt die gesamte Krümmung der Kontur von Punkt p bis zum i -ten Element dar. Die Gewichte w_i haben einen Wert

von 1 bei Punkten auf einer geraden Linie und verringern sich monoton mit zunehmender Krümmung. Für eine bestimmte einzelne Kurve Γ ergibt sich die gesamtbeachtete Auffälligkeit am Punkt p durch (8). Berechnet wird also das Maximum aller möglichen Kurven im Bild, welche am Punkt p terminieren.

$$S(p) = \max_{\Gamma} S_{\Gamma}(p) \quad (8)$$

Für praktische Berechnungen wird die Kurvenlänge der zu betrachtenden Kurven auf einen Wert N limitiert:

$$S(p) = \max_{\Gamma_N} S_{\Gamma_N}(p) , \quad (9)$$

wobei Γ_N für alle Kurven mit maximaler Länge N steht. Zur Bestimmung der auffälligen Figur müssen also alle möglichen Kurven (beschränkt auf eine Maximallänge N) durch alle Konturelemente gebildet werden und danach das Auffälligkeitsmaß berechnet werden.

3.5 Varianzen

Bei dem Ansatz mittels Varianzen geht es darum, einzelnen Bildpunkten das Attribut »auffällig« zu geben, wenn der Bildpunkt eine relativ hohe Varianz bzgl. seiner Nachbarschaft (quadratische Parzelle ungeradzahlgiger Größe) aufweist. Im Ergebnis errechnet sich das Varianzbild V gemäß (10).

$$V(x, y) = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} p(x + k - i, y + k - j) - \bar{p}^2, \quad (10)$$

mit $k = \frac{m-1}{2}$, $m = 3, 5, 7, \dots$ und \bar{p} arithmetisches Mittel. Im Anschluss daran definiert ein auf Trainingsbildern empirisch ermittelter Schwellwert oder ein Schwellwertindex (der n höchsten Varianzen) die überschwelligen Varianzen und somit die als »auffällig« zu klassifizierenden Bildpunkte.

Dieser Ansatz wird auch als einfaches Mittel zur Bewertung der Sensor- bzw. Bildqualität verwendet: Die Varianzen der Zielsignaturen werden mit denen des Hintergrunds verglichen. Sind beide Kenngrößen in etwa gleich groß, dann gilt das Bild als

»relativ schwer auswertbar«. Ein signifikanter Unterschied beider Kenngrößen deutet auf ein »relativ leicht auswertbares« Bild hin.

3.6 Dynamische Systeme

Im Allgemeinen beschreiben dynamische Systeme physikalische Prozesse mit zeitlichem Veränderungsverhalten. Iterative Ansätze zur Bildauswertung konsultieren den Bildbereich nicht nur einmal, sondern wiederholt. Die Ergebnisse der einen Auswertung parametrisieren weitere Durchläufe. Auf diese Weise kann iterativ Kontextinformation gesammelt und für den nächsten Durchlauf genutzt werden. Letztlich stellen die schon seit langem bekannten Skalenräume eine Variante iterativer Verfahren dar: die Ergebnisse der Auswertung einer Skalenebene parametrisieren die Auswertung der nächsten.

Eine der theoretischen Grundlagen für diese Ansätze liefert die in der Bildauswertungsgemeinde selten erwähnte Stabilitätstheorie von A. M. LJAPUNOV (1857-1918), welcher in seiner Dissertation 1892 dynamische (»iterative«) Systeme $[\mathcal{S}, \mathbf{r}^n]$ betrachtete:

Definition: Ein dynamisches System $[\mathcal{S}, \mathbf{r}^n]$ besitzt einen Zustandsraum \mathcal{S} und eine Funktion $\mathbf{r}^n(\mathbf{x})$, welche die (zeitliche) Entwicklung der Zustände $\mathbf{x} \in \mathcal{S}$ beschreibt, den der Zustand \mathbf{x} zum Zeitpunkt n einnimmt.

Der Zustand \mathbf{s} ist im Gleichgewicht, falls sich \mathbf{s} zeitlich nicht mehr ändert, er heißt stabil, falls kleine Änderungen bei ihm keinen beträchtlichen Einfluss auf dessen zeitliche Entwicklung nimmt und er heißt asymptotisch stabil, wenn sich der Zustand nach einer Änderung zeitlich-asymptotisch wieder dem Ausgangszustand annähert. Asymptotische Gleichgewichtszustände werden in der Chaos-Theorie sowie im Bereich der rückgekoppelten Neuronalen Netze (z.B. Hopfield-Netze, Boltzmann-Maschinen) auch Attraktoren genannt. Diese Zustände sind erwünscht, wenn es sich um Minimierungsprobleme rankt:

DEFINITION: Sei $[\mathcal{S}, \mathbf{r}^n]$ ein dynamisches System. Eine Funktion $E(\mathcal{S})$, mit $\mathcal{S} \rightarrow \mathbb{R}$ heißt Ljapunovfunktion für das dynamische System $\Leftrightarrow E(\mathbf{r}^n(\mathbf{s}))$ mit wachsendem n

monoton fallend ist. Für beliebige Zustände \mathbf{s} und n, n' , mit $n' > n$ gilt dann $E(\mathbf{r}^{n'}(\mathbf{s})) \leq E(\mathbf{r}^n(\mathbf{s}))$.

Eine solche Ljapunovfunktion wird in der physikalisch orientierten Literatur zumeist »Energiefunktion« (E) genannt. Wesentlich ist noch die Betrachtung der Stabilität des Systems:

STABILITÄTSTHEOREM VON LJAPUNOV: Sei E eine Ljapunovfunktion für $[\mathbf{S}, \mathbf{r}^n]$. Wenn E für \mathbf{s} ein lokales Minimum besitzt, dann ist \mathbf{s} stabil.

Die Stabilitätstheorie spielt eine wesentliche Rolle bei diversen Ansätzen, wie z.B. Lernstrategien, den Markov-Ketten, Minimierungen von Energiefunktionen, dem Ising-Modell, bis hin zu Funktionalen, wie dem Mumford-Shah-Ansatz.

3.7 Segmentierung mit Funktionalen

Falls kein explizites Zielobjektmodell spezifiziert werden kann (oder soll), eignen sich Funktionale zur Einbeziehung von Information in den Detektionsprozess. In diesen Fällen werden globale Informationen über die im Bild zu detektierenden Modelleigenschaften in Form von Bedingungen (*constraints*) artikuliert. Die Bedingungen werden als zu minimierendes Funktional ausformuliert. Ein solches Funktional besteht aus einem Term zur Berechnung der Ähnlichkeit von Modell- und Bildmerkmalen (und damit einer Art Beschränkung, die das Modell zwingt, so ähnlich wie möglich zu sein) sowie einem Term zur graduellen Anpassung des Modells (und damit einer langsamen, kontinuierlichen Veränderung des Modells). Zielobjektsignaturen werden sich aufgrund irgendwelcher Diskontinuitäten vom Bildhintergrund abheben (Kontrast, Farbe, Textur, ...). Es macht also Sinn, diese Übergänge gesondert zu behandeln. Eine einfache Ähnlichkeitsfunktion ist gegeben durch $\|u - u_0\|_2$, mit u zu berechnendes Merkmal (Bild, Detektions- oder Segmentierungsergebnis usw.) und u_0 Ausgangsbild bzw. -merkmal. Das Integral der Ableitung zwischen beiden wird mit der L_2 -Norm berechnet, indem u angepasst wird. Ein Beispiel eines Terms zur graduellen Anpassung realisiert

$$|\nabla u|^2, \tag{11}$$

mit ggf. der Einschränkung auf Bildbereiche ohne Berücksichtigung der Diskontinuitäten. Diese werden mit einem weiteren Term dadurch ins Spiel gebracht, indem man den Beitrag der Diskontinuitäten vom Gesamtbild berechnet. Beim Mumford-Shah-Ansatz (Mumford, et al., 1989) geht es um die iterative Minimierung einer Energiefunktion. Der Ansatz eignet sich für:

- die Segmentierung von Bildern
- die Bildverbesserung und
- Superresolution.

Gegeben sei ein initiales Bild u_0 einer Domäne Ω . Wir suchen ein »besseres«, vorliegend also segmentiertes Bild u mit korrespondierender Kantenmenge Γ durch Minimierung der folgenden Energiefunktion:

$$E[u, \Gamma | u_0] = \alpha \int_{\Gamma} ds + \beta \int_{\Omega} |u - u_0|^2 dx + \gamma \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx, \quad (12)$$

mit s Kantenlänge und α, β, γ freie positive Parameter. Dabei gilt:

- Das Ergebnisbild u stellt eine Approximation des Ausgangsbildes u_0 dar. In der Literatur wird bezüglich u oft von einem »Cartoon«-Bild gesprochen, welches die »wesentlichen« Bildinformationen von u_0 erhalten soll.
- Je kleiner E wird, desto besser wird u_0 durch $[u, \Gamma | u_0]$ segmentiert.
- Die drei verwendeten Terme stellen Kostenfunktionen bei der Optimierungsaufgabe dar:
 - $\alpha \int_{\Gamma} ds$ berechnet die Kantenlänge und entspricht dem eindimensionalen Hausdorff-Maß $\alpha \mathcal{H}^1(s)$.
 - $\beta \int_{\Omega} |u - u_0|^2 dx$ vergleicht u mit dem Ausgangsbild u_0 .
 - $\gamma \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx$ berechnet die Weichheit von u ohne die Kanten zu berücksichtigen.
- Die Optimierungsaufgabe kann z.B. mittels Stimulated Annealing (Geman, et al., 1984), Splines (Cremers, et al., 2002), Graduated Non Convexity (GNC) (Grady, et al., 2008), Level Set Methoden (Tsai, et al., 2001), (Vese,

et al., 2002), Relaxation (Pock, et al., 2009) und anderen Ansätzen durchgeführt werden. Grundsätzlich ist die effiziente Lösung dieser Optimierungsaufgabe nach wie vor Gegenstand der Forschung.

3.7.1 Detektion mittels CFAR-Ansätzen

Das Ziel der Detektion (Entdeckung von ROI (siehe Einführung) bzw. »Points of Interest« (POI)) ist die Datenreduktion (Reduktion auf relevante Bildpunkte). Die meisten Ansätze verwenden die Bildpunktintensitäten bzw. deren Verteilung als primäres Merkmal für den Detektionsschritt. Im Wesentlichen werden 3 Methoden unterschieden:

1. histogrammbasiert
2. kantenbasiert und
3. musterbasiert.

Die musterbasierte Methodik verwendet sehr häufig die Korrelation. Ein Vergleichsmuster (Referenz) wird über den Bildbereich verschoben und für jeden abgedeckten Musterbereich im Bild die Korrelation berechnet. Im Resultat erhält man ein neues Bild (Korrelationswertbild). Auf dieses wird dann zwecks Maximasuche ein Schwellwertverfahren angewendet.

CFAR (engl. *Constant False Alarm Rejection*) Ansätze konzentrieren sich auf Bildbereiche, deren Größe von der erwarteten Zielobjektgröße abhängen. Bildpunkte, deren Intensitäten herausragen, werden als »Zielobjekt« deklariert. Daher kann ein einzelnes Zielobjekt mehrere CFAR-Detektionen anziehen. Die CFAR-Detektionen werden zusammengefasst (engl. *Clustering*) und ein CFAR-Bild wird generiert. Für jede Testzelle werden folgende Berechnungen angestellt:

$$\text{Objekt} = \frac{I - \mu_c}{\sigma_c} > k_{CFAR}, \quad (13)$$

mit I Intensität des betrachteten Bildpunktes, μ_c geschätzter Mittelwert der Clutterintensität, σ_c geschätzte Standardabweichung der Clutterintensitäten sowie k_{CFAR} konstanter Schwellwert. Bis auf die Intensität werden alle weiteren Parameter aus Lernstichproben gewonnen. Jeder Bildpunkt, welcher überschwellig ist ($>k_{CFAR}$) gilt

als zugehörig zu einem Zielobjekt. Alle unterschwelligen Bildpunkte werden dem Hintergrund (bzw. Clutter) zugeordnet.

3.8 Hintergrundcharakterisierung

Die Detektion von Zielobjektsignaturen in Bilddaten stellt im Wesentlichen eine Zweiklassenklassifikation dar. Unterschieden werden soll zwischen den Klassen »Zielobjekt« und »Hintergrund« im Bild. Es ist evident, dass diese Aufgabe bei kontrastreichen Zielsignaturen vor homogenem Hintergrund trivial ist. In der vorliegenden Anwendungsdomäne liegt jedoch alles andere als der triviale Fall vor: die gegebenen Bildfiguren sind schwer zu deuten und hier befinden sich die Detektionsrate (P_D – *Probability of Detection*) und die Falschalarmrate (FAR – *False Alarm Rate*) in einem kanonisch konjugierten Zusammenhang: je besser der eine Wert, desto schlechter der andere. Eine hohe Detektionsrate ist einfach zu erreichen, allerdings unter Inkaufnahme einer ggf. sehr ungünstigen Falschalarmrate. Die Ursache für einen generierten Falschalarm liegt auf der Hand: eine der Klasse »Hintergrund« zuzuordnende Bildfigur wies eine hinreichende Ähnlichkeit zum verwendeten Zielobjektmodell des Detektors auf. Demgegenüber resultieren »gutartige« (z.B. homogene) Hintergrundbereiche nicht in Falschalarmen. Der Bildhintergrund beeinflusst jedoch nicht nur die Leistungsparameter des Detektors, sondern auch den Ansatz seiner Implementierung. Es ergeben sich folgende Schlussfolgerungen:

1. Im Gegensatz zur klassischen Vorstellung, sich »nur um die Zielobjektsignaturen« zu kümmern (u.a. nur diese zu trainieren etc.), muss auch bestimmten Hintergrundfiguren Aufmerksamkeit zu teil werden.
2. In der Klasse »Hintergrund« sollte es eine weitere Klasse von solchen Bildfiguren geben, welche als »falschalarmkritisch«, da ähnlich zur Zielsignatur, eingestuft wird. Diese Klasse wird im Folgenden »Clutter« genannt.
3. Die Klasse »Clutter« verdient wie die Klasse »Zielobjekt« eine besondere Aufmerksamkeit. Während man durch die Analyse der Zielobjektsignaturen die gewünschte Detektionsrate sicherstellen kann, kann die Analyse der Clutterstrukturen die Falschalarmrate senken.

Der Hintergrund, insbesondere in Form von Clutter, reduziert die Systemleistung im Hinblick auf die Detektionsaufgabe. Das Ziel der Detektion ist die Unterscheidung zwischen Zielobjekt und Hintergrund. Daher wird in semantischer Hinsicht die Sensorantwort auf ein Szenenzielobjekt oft als »Signal« und die Sensorantwort auf den Hintergrund (insbesondere Clutter) als »Rauschen« bezeichnet. Da insbesondere der Clutter die größten Probleme im Hinblick auf die Detektion bzw. die Senkung der Falschalarmrate darstellt, wird wiederum im semantischen Sinne zumeist von einer »Signal-/Clutter-Ratio« (SCR) gesprochen. Je kleiner dieses Verhältnis ist, desto schwieriger ist es, das gegebene Bild auszuwerten. Die SCR ist also auch eine Möglichkeit zur Bewertung der Bildqualität bzw. zur Schätzung von Leistungsparametern eines maschinellen oder menschlichen Auswerters bei einer Detektionsaufgabe. Ein häufig benutztes Cluttermodell C nutzt die Varianz der Intensitäten:

$$C = \sqrt{\frac{1}{N} \sum_{i=1}^N \sigma_i^2}. \quad (14)$$

Dabei wird das Szenenbild in N quadratische, aneinander angrenzende Fenster (bzw. Kacheln) aufgeteilt, wobei jedes Fenster die doppelte Größe einer Zielsignatur hat. Der Clutter wird somit durch die Quadratwurzel der normierten Summe der Varianzen aller Fenster i berechnet. Zur Berechnung der SCR wird ein Fenster mit enthaltener Zielsignatur gewählt, die Varianz berechnet und durch den Wert C dividiert. Die SCR kann also bei diesem Ansatz nur über die Kenntnis der Grundwahrheit (engl. *ground-truth*) berechnet werden, wohingegen das Cluttermodell C ohne die Grundwahrheit auskommt.

Die vorgestellte Methode zur Berechnung einer SCR kann an einigen Stellen verfeinert werden. Zum einen, indem das Bild nicht in aneinander angrenzende Fenster aufgeteilt wird, sondern – ähnlich wie bei Filteraufgaben – eine Berechnung mittels dicht überlappender Fenster erfolgt. Ferner ist es zur Berechnung der SCR notwendig, jene Fenster aus der Berechnung von C auszuschließen, die die Zielsignatur vollständig oder in Teilen enthalten, da ansonsten Teile von Zielsignaturen als Clutter in die Berechnung von C bzw. SCR einfließen. Außerdem kann man das SCR-Modell leicht auf die Präsenz mehrerer Zielsignaturen erweitern.

Die getätigten Überlegungen weisen auf die in der Literatur beschriebenen generellen Ansätze zur Bildbewertung bzw. Schätzung von Leistungsparametern von Mensch oder Maschine hin. Gesucht wird dabei eine Maßzahl C , die möglichst in linearem Zusammenhang mit der ggf. vorauszusagenden Detektionswahrscheinlichkeit P_D eines Systems steht:

$$P_D = k \cdot C \quad (15)$$

mit k Konstante. Die vielen unterschiedlichen Ansätze unterscheiden sich zum einen darin, dass sie mit oder ohne Grundwahrheit operieren und ansonsten im Wesentlichen nur noch durch die Anzahl und den Typ der berechneten Merkmale:

1. Bestimmung eines oder mehrerer als »aussagekräftig« angesehenen Merkmale (im obigen Ansatz war es das Merkmal »Varianz«).
2. Bildlokale Berechnung der Merkmale und Zusammenfassung via Summenbildung zu einer bildglobalen, als »charakteristisch« angesehenen Maßzahl C , die als »Bildgüte« interpretiert wird. C berechnet sich nach (16), wobei gilt: f Normierungsfunktion, M Merkmale, C_i bildglobale Cluttermaßzahl bzgl. Merkmal i (z.B. Intensität gemäß (14)) und g_i (verschiedene) Normierungsfunktionen g_i der einzelnen Cluttermaßzahl. Im Ansatz steht f für Quadratwurzel, g für Normierung ($1/N$) und da nur ein Merkmal berechnet wird (die Varianz), ist $M=1$:

$$C = f \sum_{i=1}^M g_i(C_i) . \quad (16)$$

3. Falls die Grundwahrheit hinzugezogen wird, werden die obigen Berechnungen auf die Bildbereiche angewendet, in denen Zielsignaturen enthalten sind und durch C dividiert. Somit erhält man eine SCR.

Kritik

Es liegt auf der Hand, dass bei einer Cluttermaßzahl C (mit oder ohne Zuhilfenahme der Grundwahrheit), die SCR nur dann relevant ist, wenn die Merkmale berücksichtigt werden, die auch von einem maschinellen System bzw. vom Menschen erfasst werden. Beispielsweise wird durch die ausschließliche Berechnung der Varianz für C

bzw. SCR unterstellt, dass die Grauwertvarianz im Bild das entscheidende Unterscheidungsmerkmal von Zielsignatur und Hintergrund (bzw. Clutter) ist. Zum einen ist dem keineswegs immer so und zum anderen wäre damit die ATR-Aufgabe gelöst, denn wäre die Varianz hinreichend trennungsrelevant, dann würde es reichen den Detektionsalgorithmus lediglich die Grauwertvarianz aller möglichen Bildparzellen ausrechnen zu lassen und entsprechende Parzellen mit überschwelliger Varianz als Detektionshypothesen vorzuschlagen. Im Grunde wird also durch diese Ansätze nichts anderes gemacht, als in »abgespeckter« Form ein vereinfachtes, stark heuristisches ATR-Verfahren zu implementieren und zu hoffen, dass die vereinfachte ATR-Variante mit seinen Leistungsparametern einem ausgeklügelten ATR-Verfahren entspricht.

Dass z.B. die Berechnung der Varianz keineswegs ausreicht, um Zielsignaturen im Bild zu finden, soll Abbildung 13 verdeutlichen. Dargestellt sind Dreiecke, Kreise und Rechtecke. Die Aufgabe ist es, Rechtecke zu detektieren. Die Cluttermaßzahl auf der Basis der Varianz zeigt ein »schwieriges« Bild, denn die Varianzen der Parzellen, in denen sich Dreiecke oder Kreise befinden, unterscheiden sich nicht von der Parzellenvarianz mit dem enthaltenen schwarzen Rechteck. Ferner ist die Varianz für die Parzelle, in der das graue Rechteck ist, sehr niedrig. Die Varianz als Merkmal wäre nicht charakteristisch und auch nicht trennungsrelevant, um Clutter von Zielsignatur zu unterscheiden. Im Ergebnis würde eine zu niedrige Detektionsrate und zu hohe Falschalarmrate prognostiziert. Ein ATR-Verfahren, das als Merkmal die Form der Zielsignaturen berücksichtigt, hätte mit dem gegebenen Bild keine Schwierigkeiten.

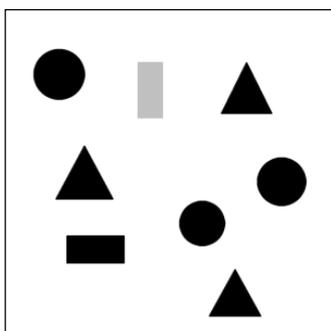


Abbildung 13: Beispielbild »Detektion von Rechtecken«.

3.9 Texturbasierte Ansätze

Texturbasierte Ansätze sind in Bildern günstig, die einen hohen Anteil natürlicher und großflächiger Hintergrundstrukturen aufweisen (z.B. Waldflächen). Demgegenüber sind direkte texturbasierte Segmentierungstechniken von Zielobjekten (z.B. Fahrzeugen) meist sehr heuristischer Natur, da Zielobjekte relativ klein sind und daher keine ausgeprägte Textur besitzen. Indirekte Methoden können jedoch durchaus erfolgreich sein: Zunächst erfolgt eine großflächige, relativ grobe texturbasierte Segmentation des Bildes. In jedem segmentierten Bildbereich wird dann nach »Texturfehlern« gesucht. D.h. Stellen, die dem Homogenitätskriterium des jeweiligen Texturbereichs nicht genügen. Diese Stellen werden dann als Zielobjekt angesehen. Texturcharakteristika werden z.B. mittels Grauwertübergangsmatrizen ermittelt. Andere Ansätze ermitteln einen Texturgradienten, der zur Betrachtung menschlicher Tiefen- und Größenwahrnehmung eingeführt wurde (Gibson, 1950). Texturgradienten können neben der Bildsegmentierung auch für einfache Tiefenschätzungen verwendet werden, denn weiter entfernt liegende Bildfiguren erscheinen dichter gepackt (vgl. Abbildung 14). Der Eindruck dieser vermeintlich dichteren Packung wird auch bei der »Ponzo-Größenillusion« ausgenutzt (Renier, et al., 2005). Ein Texturgradient lässt sich für jeden Bildbereich (in 3D) z.B. wie folgt ermitteln:

$$\text{Texturgradient} = \sqrt{\mu_i - \mu_{i+3}^2 + \sigma_i - \sigma_{i+3}^2}, \quad \text{für } 0 < i \leq 3. \quad (17)$$

Aus diesen Werten wird ein Texturgradientenbild erzeugt, aus welchem ein Texturgradientenhistogramm generiert wird. Das Histogramm wird anschließend analysiert, um die Bereiche für die Textursegmentation (Hintergründe, Zielobjekt) festzulegen.

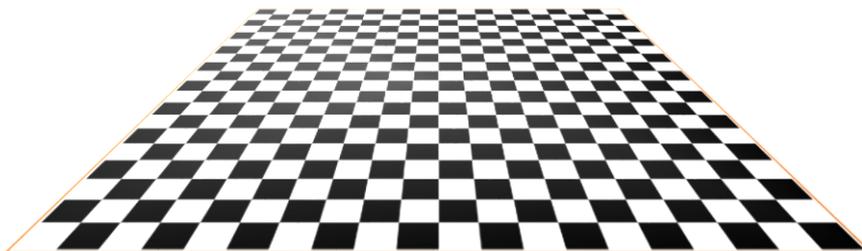


Abbildung 14: Demonstration eines visuell wahrnehmbaren Texturgradienten.

Bei den Relaxationsmethoden (vgl. z.B. (Hansen, et al., 1997)) zur Segmentierung wird für jeden Bildpunkt die Wahrscheinlichkeit geschätzt, ob er zum Zielobjekt oder zum Hintergrund gehört. Dieser Schätzwert basiert auf den Distanzen vom Ende und dem Mittelwert der Grauwertskala. Die Wahrscheinlichkeiten werden iterativ aktualisiert, basierend auf den Wahrscheinlichkeiten von benachbarten Bildpunkten.

3.10 NAG-Auffälligkeitsdetektion

Die Lösungsansätze in (Korn, 1988) bzw. (Korn, 1995) wurden in der Folge verschiedentlich verwendet (z.B. (Grimm, 2003)) und können als artverwandte Vorläufer des heute weit verbreiteten SIFT-Ansatzes (Lowe, 2004) angesehen werden. In beiden Fällen werden Merkmale aus dem Bildbereich berechnet und insbesondere Gradientenrichtungen in Histogrammen zusammengefasst. Während KORN Histogramme für Bildparzellen gemäß der vermuteten Zielobjektgröße erstellt und analysiert, erstellt LOWE auf der Basis lokaler Histogramme sog. »Keypoint Vectors«. Die Ansätze von KORN stehen im Anwendungskontext der Erkennung von Objekten (primär Fahrzeugsignaturen bei Senkrechtsicht) in natürlicher Umgebung, wohingegen sich LOWE »nur« auf die Extraktion möglichst invarianter Merkmale bezieht.

KORN schlug ein Verfahren zur Berechnung und Auswertung von *Gradientenrichtungshistogrammen* vor, welches als zusätzliches Entscheidungskriterium eine globale Grauwertschwelle nutzt (Korn, 1995 S. 109-112). Die Prämissen des Ansatzes, also die Definition der Auffälligkeit, stellen sich wie folgt dar:

- Parallele und/oder orthogonale Linienstrukturen.
- Senkrechtsicht der Zielsignaturen (ansonsten Verletzung des Orthogonalitätskriteriums).
- Hinreichend kontraststarke Zielsignaturen.
- Homogene Bereiche innerhalb der Zielsignaturen, um sie bei einer Binarisierung als Flecken ballen zu können.

- Objekte befinden sich in natürlicher Umgebung, d.h. im Gelände und nicht
 - in Siedlungsgebieten
 - auf Straßen
 - verursachen keine Spuren,
 da es ansonsten zwangsläufig zu Falschalarmen kommt.
- Die Objektsignatur ist vollständig sichtbar, sie ist nicht verdeckt oder verläuft an einer Seite kontinuierlich mit dem Hintergrund.
- Die Objektsignatur ist in der relevanten Bildparzelle kontrastreicher, d.h. die Grauwertdifferenz ist größer, als andere Strukturen im Parzellenbereich.

Die ideale Fahrzeugsignatur vor idealem Hintergrund gemäß obiger Auffälligkeitsprämissen sieht wie folgt aus:

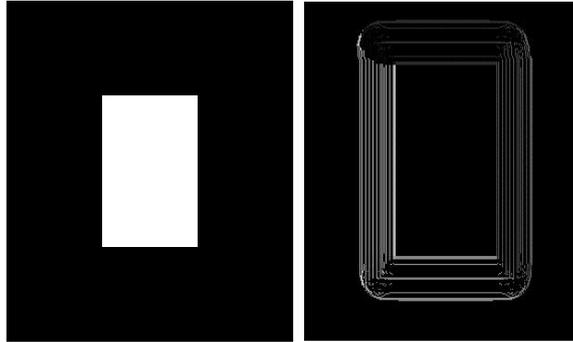


Abbildung 15: Ideale Fahrzeugsignatur gemäß Korn'schem Ansatz links und vergrößerte Gradientenrichtungen (grauwertcodiert) rechts.

Bei diesem Verfahren wird das Eingabebild f zunächst mit einem Gradientenoperator (normierte Ableitung einer Gaußfunktion g , NAG genannt) gefaltet (wir betrachten aufgrund der Separierbarkeit der Gaußfunktionen sowie ihrer Ableitungen nur die eindimensionale Variante); eine neue Herleitungsvariante des NAG sowie die Beleuchtung einiger Eigenschaften dieser Faltungsfunktion ist im Abschnitt 8 dargestellt. Es wurde aus Gründen der Übersichtlichkeit folgende Gaußfunktion zugrunde gelegt, mit σ als Standardabweichung (präziser: $\sigma / \sqrt{2}$, zur besseren Übersichtlichkeit wird $\sqrt{2}$ weggelassen):

$$g_{\sigma} = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2}{2\sigma^2}}. \quad (18)$$

Die Ableitung von g ergibt sich zu:

$$g'_\sigma = -\frac{x}{2\pi\sigma^4} e^{-\frac{x^2}{2\sigma^2}} \quad (19)$$

oder vereinfacht geschrieben:

$$g'_\sigma = -\frac{x}{\sigma^2} g_\sigma(x) \quad (20)$$

KORN führte hier einen Normierungsfaktor $k = \sigma\sqrt{2\pi}$ ein, der eine Skalierungsinvarianz realisiert und damit den Vergleich von Ergebnisbildern für verschiedene Auflösungen ermöglicht (Korn, 1995):

$$g'_\sigma = -\frac{k \cdot x}{\sigma^2} g_\sigma(x). \quad (21)$$

Aus den Faltungsergebnissen lassen sich die Amplituden und Richtungen der Gradienten berechnen (Abschnitt 9.1). Anschließend werden nur noch die Gradientenrichtungen berücksichtigt, deren Amplitude über einem Schwellwert m liegt (meist $m=10$ (empirisch)), um Rauscheinflüsse zu mindern. Nachdem eine Kachelung (Eingangsgröße s gemäß zu erwartender Zielsignaturgröße) des Grauwertbildes mit 50%-Überlappung durchgeführt wurde, splittet sich das Verfahren auf: Für jede Kachel wird zum einen das korrespondierende Gradientenrichtungshistogramm erstellt und zum anderen ein Kontrastmaß berechnet.

Als Kontrastspannweite K kann in Anlehnung an Kontrastdefinitionen der Psychophysik (wie bereits dargelegt) nach KORN z.B. verwendet werden:

$$K = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}}, \quad (22)$$

wobei L_{\max} bzw. L_{\min} die maximale bzw. die minimale Leuchtdichte (Pixelintensität) in der betrachteten Kachel des Eingangsbildes f bezeichnen. Als Spannweite S einer Kachel bezeichnete KORN die Differenz des maximalen und minimalen Intensitätswerts in der betrachteten Kachel:

$$S = L_{\max} - L_{\min}. \quad (23)$$

Aus den Spannweiten aller Bildkacheln wird aus der mittleren Spannweite m_s und deren Standardabweichung σ_s über die Formel

$$G_s = m_s + h \cdot \sigma_s \quad (24)$$

eine adaptive und bildglobale Spannweiteschwelle G_s berechnet. Die Konstante h wird empirisch aus einer Lernstichprobe gewonnen (Praxiswerte schwanken meist um den Wert 1,5).

Die Auswertung der Richtungshistogramme in den Kacheln dient der Feststellung, ob Vorzugsrichtungen in bestimmten Abständen präsent sind. Parallele Kanten resultieren in Maxima des Richtungshistogramms im Abstand von 180° ; orthogonale Kanten in Maxima im Abstand von 90° . Ein gut ausgeprägtes Rechteck besitzt also 4 signifikante Maxima im Abstand von 90° (vgl. Abbildung 16).

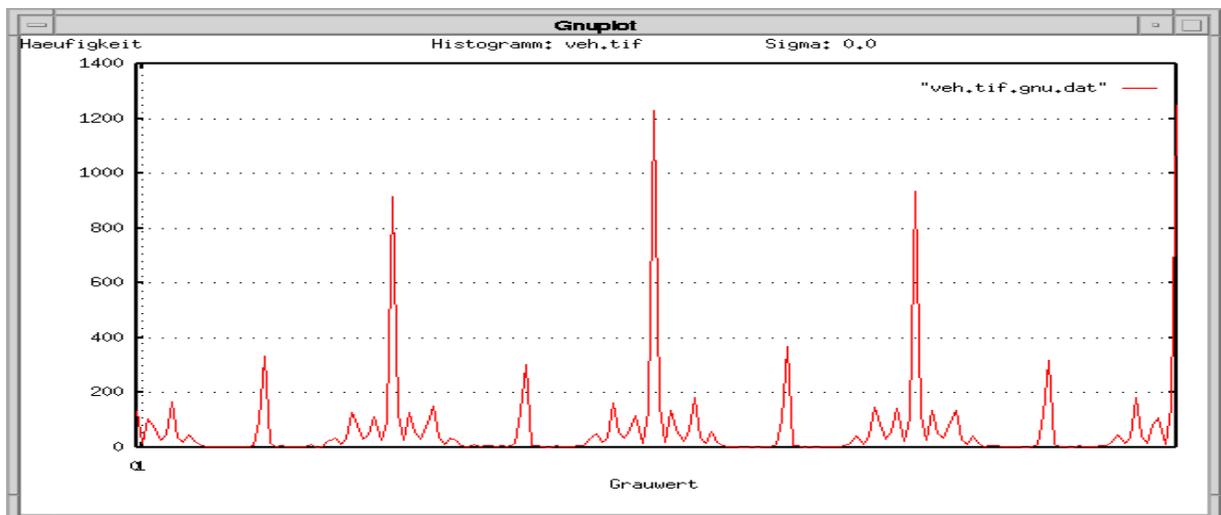
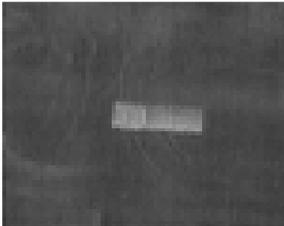


Abbildung 16: Senkrechtsicht auf ein Fahrzeug und korrespondierendes Gradientenrichtungshistogramm (zu beachten: das vierte signifikante Maximum liegt auf der rechten Ordinate).

Zur Detektion möglicher Maxima wendete KORN ein einfaches Verfahren an: Das Histogramm wird mit einem Gauß-Tiefpass geglättet und ausgehend vom höchsten

Wert im Histogramm (= globales Maximum) wird im Abstand von 90° ($\pm 5^\circ$ - 16°) nach lokalen Maxima gesucht. Ein lokales Maximum liegt vor, wenn es in dem entsprechenden Winkelintervall mindestens einen Wert gibt, der größer ist als sein linker und sein rechter Nachbar.

Als Entscheidungskriterium für das Attribut »auffällig« einer Kachel (bzw. Rasterelement) gibt (Korn, 1995 S. 112) an:

»Ein Rasterelement ist auffällig, wenn seine Spannweite größer als $(m_s + 1,5 \cdot \sigma_s)$ ist **und** signifikante Maxima des Richtungshistogramms einen Abstand von (90 ± 16) Grad haben.«

Ein Programmablaufplan ist in Abbildung 17 zu sehen.

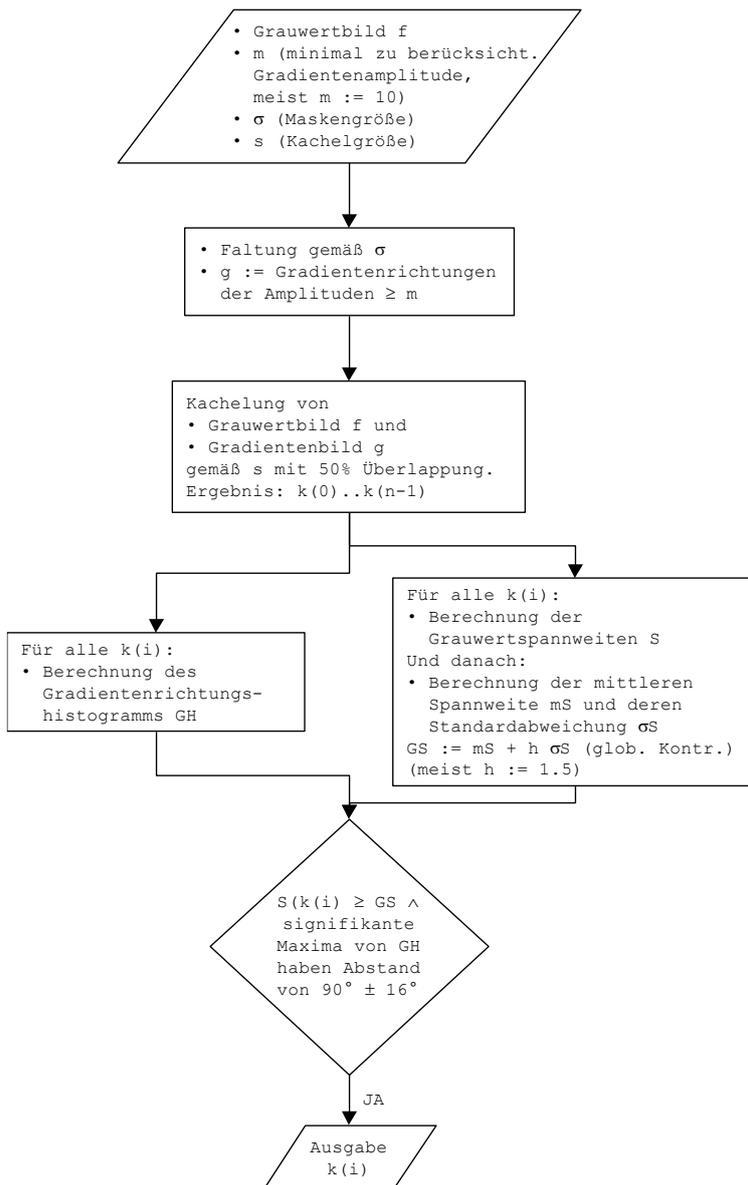


Abbildung 17: Programmablaufplan »ROI-Screening«.

3.10.1 Eigene Bewertung des Korn'schen Verfahrens

Das vorgestellte Verfahren lässt sich effizient implementieren. Insbesondere die grobe Kachelung ermöglicht Realisierungen mit geringem Rechenaufwand. Allerdings ist es in vielen Fällen in der vorgeschlagenen Form nicht mit hinreichend guten Resultaten einsetzbar und es erfordert einen recht hohen empirischen Aufwand für die Parameterfindung, der nicht von einem unerfahrenen Anwender vorgenommen werden kann.

Die Nachteile beginnen mit der Einrichtung des Verfahrens in der Trainingsphase, denn diese kann nur durch einen kundigen Experten, nicht jedoch durch einen Anwender erfolgen und ist ggf. zeit- und arbeitsintensiv, da das Verfahren über zu viele freie Parameter verfügt. Die Probleme resultieren u.a. aus diesen freien Parametern bzw. Schwellwerten, die durch ein zeitaufwendiges Training ermittelt werden müssen:

- Festzulegen ist z.B. der Schwellwert h , der die globale Kontrastschwelle festlegt, ab wann Zielobjekthypothesen zu betrachten sind und damit eine Kontrastabhängigkeit impliziert, womit die ursprünglich gegebene, weitgehende Kontrastunabhängigkeit in Teilen verloren geht. Der Parameter h soll Rausch- bzw. Cluttereffekte vermeiden helfen und somit die FAR reduzieren.
- Zu bestimmen ist auch der Schwellwert m , der die minimale Länge der Gradienten definiert, deren Gradientenrichtungen bei der Histogrammerstellung berücksichtigt werden sollen. Der Parameter m soll Rauscheffekte unterdrücken. Durch die starre Schwelle reduziert man zwar Falschalarme, jedoch werden auch hier kontrastschwache Objekte (einer der Hauptvorteile der statistischen Auswertung von Gradientenrichtungen) nicht mehr entdeckt. Auch dann nicht, wenn sich die kontrastschwachen Objekte ansonsten idealtypisch präsentieren.
- Auch der Skalierungsparameter σ für die Implementierung des NAG wäre zu bestimmen.
- Der Faktor 1,5 in der Formel $m_s + 1,5 \cdot \sigma_s$ ist rein empirisch ermittelt. Auch er trägt als zweites zu erfüllendes Kriterium für »Auffälligkeit« zu einer weiteren Verschärfung der Kontrastabhängigkeit bei.
- Die Kachelüberlappung von 50% gewährleistet zwar effiziente Realisierungen und eine per se reduzierte Falschalarmrate, jedoch bekommt man auf diese Weise keine guten Detektionsresultate bei Zielobjektsignaturen, die auf den Kanten der Kachelung liegen. Der Überlappungsgrad wäre ein einzustellender Parameter, der anhand von Lernstichproben zu bestimmen wäre.
- Die Bestimmung der Histogrammmaxima ist in manchen Fällen unzuverlässig. Die signifikanten Maxima des Richtungshistogramms sollen einen Abstand von (90 ± 16) Grad haben. Die Spannweite $\pm 16^\circ$ ist rein empirisch.

Auch die Bestimmung der Maxima ist ggf. problematisch: es wird nach Wertdifferenzen in der genannten Umgebung geschaut und dann im Abstand von 90° dieser Vorgang wiederholt. Dabei kann man signifikante Zwischenmaxima übersehen oder auch ein kleines lokales Maximum als »Auffälligkeitsmaximum« ansehen. Die Glättung des Histogramms erleichtert zwar die (problematische) Maximabestimmung, jedoch verändert sie auch die Statistik der Gradientenrichtungen und verbessert auch nicht die Qualität der Erkennung.

- Die Auswertung der Richtungshistogramme hat zwei große Vorteile: a) Prinzipielle Kontrastunabhängigkeit sowie b) Richtungsinvarianz. Abgesehen von der Einschränkung der Kontrastunabhängigkeit durch die limitierenden Parameter h und m geht der Ortsbezug verloren denn das Histogramm speichert nicht die Information, *wo* in der Kachel die entsprechenden Gradientenrichtungen vorkommen, und somit ist auch unbekannt, in *welcher* strukturellen Anordnung sich die Gradienten zueinander befinden.
- Die Maximumdetektion im Histogramm der Gradientenrichtungen ist nicht erläutert. Dabei ist die Maximumdetektion nicht trivial. Schnell können durch Rauscheffekte und sonstige Einflüsse »zufällige« Lokalmaxima das Verfahren in die Irre führen. Neben ad-hoc Implementierungen (Suche nach größten Werten in gewisser Umgebung) gibt es auch bei herkömmlichen Standard-Maximadektoren (glätten, differenzieren, Nulldurchgangsdetektion) Probleme aufgrund ihrer grundsätzlichen Empfindlichkeit (Zufallseffekte). Zur exemplarischen Implementierung wurde die Maximumdetektion, die letztlich auf (Heinbach, 1988) zurückgeht, gewählt, bei der zunächst lokale Maxima nach herkömmlichem Muster detektiert werden und dann eine Überprüfung durch ein Schwellenkriterium stattfindet. Schließlich wird nach globalen Maxima gesucht. Als Schwellwertkriterium wird die Ausgeprägtheit eines Maximums herangezogen, welche als die kleinere Pegeldifferenz dL zwischen einem lokalen Maximum und den unmittelbar angrenzenden lokalen Minima definiert ist. Wenn die zwei angrenzenden lokalen Minima innerhalb des Schwellkriteriumbereichs liegen, dann spricht man von einem »globalen Maximum«.
- Für andere Modelltypen als parallele und orthogonale Strukturen sind softwaretechnische Änderungen und entsprechende Parametrierarbeiten erforderlich.

Bei der Einrichtung des Verfahrens wird nach der Maxime »wähle gute Zielsignaturen« gehandelt, wobei die Eigenschaft »gut« durch den kundigen Experten festgelegt wird. Dabei werden sogenannte »Chips« (Bildausschnitte mit Zielsignaturen sowie umgebenden Hintergrund⁷) aus den Trainingsbildern herausgeschnitten und als Trainingsmuster zugrunde gelegt.

Fehl- und Falschdetektionen für eine größere Menge an Beispielen sind in dem skizzierten Ansatz systemimmanent. Abbildung 18 zeigt ein solches Beispiel für eine »ideale« Falschdetektion (im Bild rechts). Vier Kanten mit radial kontinuierlichem Grauwertverlauf auf jeweils einer Kantenseite resultieren in nahezu idealen Gradientenrichtungshistogrammen.

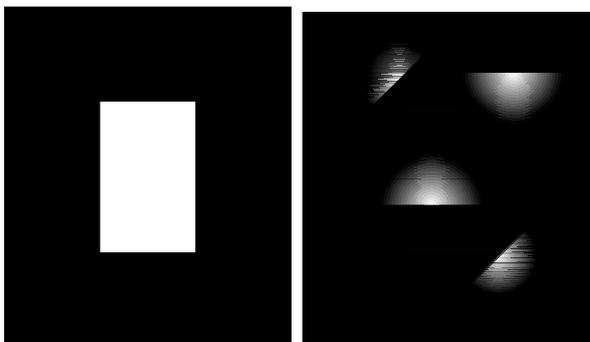


Abbildung 18: Beispiel für eine »ideale« Falschdetektion (rechts). Zum Vergleich: eine ideale Fahrzeugsignatur (links).

Auch die Annahmen, dass Fahrzeugsignaturen »heller« oder »dunkler« als der Hintergrund sein sollen, können zu Fehldetektionen führen (Abbildung 19).

⁷ Wie viel Hintergrund, ist ebenso empirisch zu testen.

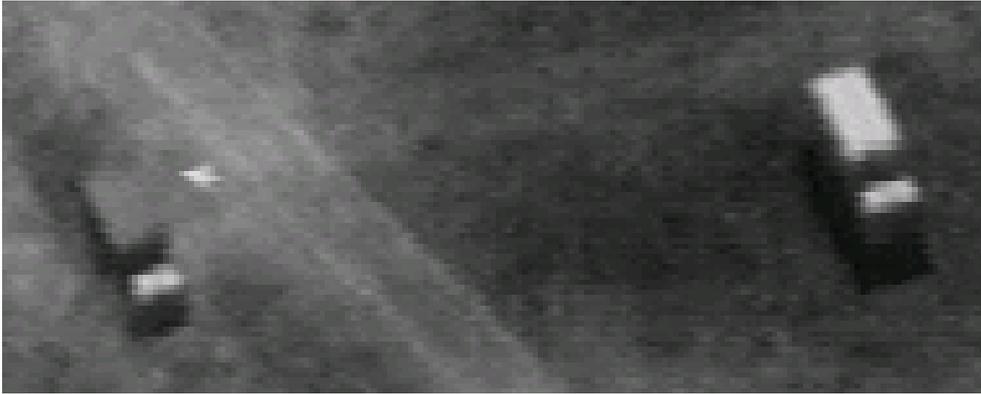


Abbildung 19: Zwei Fahrzeuge (LKW) in einem IR-Bildausschnitt. Das eine ist hell das andere dunkel. Bild: IOSB.

Zur Untersuchung der Einflussparameter wie Kontrast, Orientierung und Rauschabhängigkeit wurde vom Autor ein Testbild (eine Art »Ground-Truth«-Bild) konstruiert (vgl. Abbildung 20). Das Testbild zeigt 255 Rechtecke in den verfügbaren Grauwerten bei einem Grauwertwertebereich von 8 Bit in zufälligen Orientierungen. Zusätzlich zum Kontrastverhalten können verschiedenste Rausch-, Schmier- und Überlagerungseffekte simuliert und entsprechende Leistungsparameter von Detektionsverfahren untersucht werden. Insbesondere wird erkennbar, wie sich z.B. Kontrastschwelen und andere, kontrastsensitive Parameter eines Verfahrens auf die Detektion kontrastschwacher Objekte auswirken.

Der Hintergrund des Testbilds hat den Grauwert 128. Deshalb sind nur 255 und nicht 256 Rechtecke mit 255 unterschiedlichen Grauwerten sichtbar. Der Mensch kann auf einem Bildschirm die Rechtecke mit einer Grauwertfüllung nahe 128 nicht sehen. Das liegt einerseits am eingeschränkten Vermögen des Menschen, verschiedene Grauwerte unterscheiden zu können (je nach Quelle 60-150) und andererseits an den Einstellungen (insbes. dem Gammawert, bei Arbeitsplatzrechnern meist bei 2,2 eingestellt) bzw. Einschränkungen des Bildschirms.

Die Verarbeitung des Testbilds gibt unmittelbar Antworten auf folgende Fragen:

- Wie geht das Verfahren mit Objektsignaturen um, die nahe am Bildrand liegen (in den meisten Fällen werden dort keine Detektionshypothesen generiert)?

- Wie kontrastsensitiv ist ein eingesetztes Detektionsverfahren? Werden alle Rechtecke detektiert, die weit genug vom Rand entfernt liegen, liegt eine (maximale) »Kontrastunabhängigkeit« vor. Beispielhaft wurde auf das Testbild der diskutierte Korn-Ansatz sowie der im Zuge dieser Arbeit entwickelte Operator angewendet (Abbildung 21). Bei dem Korn-Ergebnis ist zu sehen, dass sich im Ergebnisbild in der Mitte horizontal ein leerer »Schlauch« an nicht detektierten Rechtecken bildet. Diese Fehldetektionen resultieren im Wesentlichen aus den besprochenen Kontrastschwellen. Der alternative Operator erfasst hingegen alle Rechtecke, bis auf die, die direkt an den Bildrand angrenzen.
- Mit der gewählten Ergebnisvisualisierung wird auch deutlich, dass Objekte wie in Abbildung 19 (helles Fahrzeug auf dunklem Grund, dunkles Fahrzeug auf hellem Grund) erfasst werden.
- Wie erwähnt, können beliebige Rauscheinflüsse auf die Leistungsparameter eines Detektionsverfahrens untersucht werden.

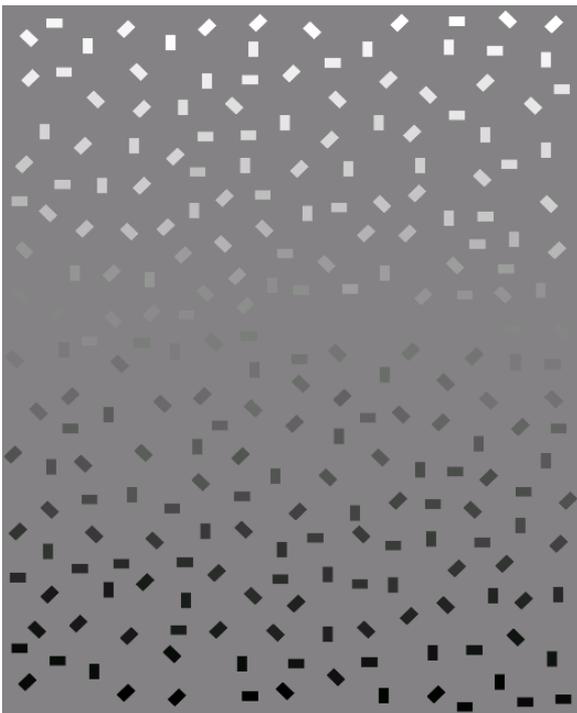


Abbildung 20: Testbild zur Überprüfung der Kontrasteinflüsse, Orientierungen und Rauscheinflüsse.

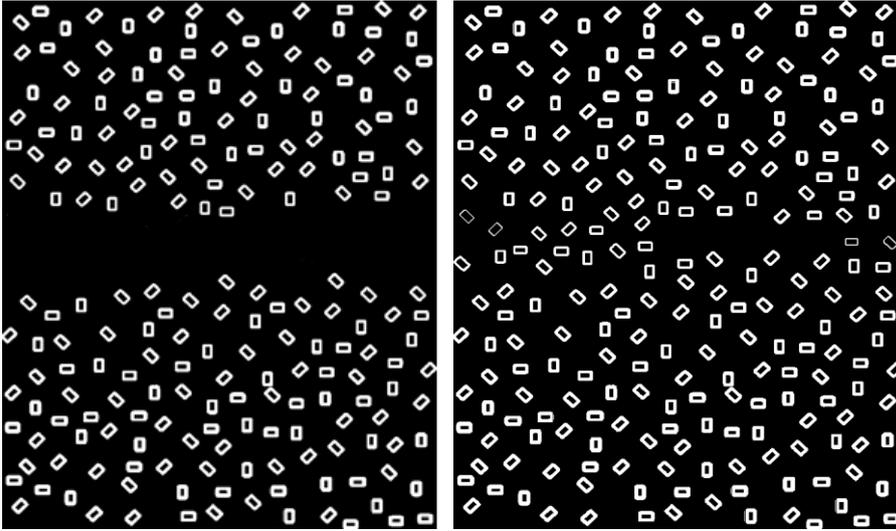


Abbildung 21: Beispielhafte Anwendung des Korn-Operators (links) und des selbst entwickelten Operators (rechts) auf das Testbild.

4 Einordnung und signalnaher Ansatz

Der signalnahe Ansatz bezieht sich auf folgende Aspekte:

1. Finden global auffälliger Strukturen: Dazu sind entsprechende Auffälligkeitsmaße für Einzelbilder zu definieren. Es soll zwischen drei Phänomenen unterschieden werden:
 - a. Primärartefakt/Auffälligkeit
 - b. Sekundärartefakt/Auffälligkeit
 - c. Hintergrundcharakterisierung
2. In Ansätzen: Propagation gefundener Auffälligkeiten (Sekundärartefakte), um nach Primärartefakten zu suchen.

4.1 Prozedur

Die in Abschnitt 3.10.1 diskutierten Nachteile, insbesondere, was die Starrheit der Modelle, diverse empirisch zu ermittelnde Parameter, Abkehr von der umfassenden Kontrastunabhängigkeit und der Maximasuche anbelangt, ist hier das Ziel, ein Verfahren vorzustellen, welches über folgende Eigenschaften verfügt. Basierend auf frühen Arbeiten (Müller, et al., 1999) werden diverse Erweiterungen vorgenommen:

- Detektion von Auffälligkeiten in einem Bild gemäß definierbarer Auffälligkeitsmodelle.
- Ein vom Ansatz her schnelles Verfahren.
- Möglichst einfache, dennoch flexible Modellannahmen.

Wenn man z.B. die Modellvorstellung für ein Auffälligkeits- bzw. Salienzmaß hat, dass sich die gesuchten MMO durch Parallelität und Orthogonalität auszeichnen, so kann man mit dem hier beschriebenen Ansatz entsprechende Salienzmodelle a priori vorgeben und benötigt keine umfangreiche Trainings- bzw. Einrichtphase. Dabei ist der gewählte Ansatz obendrein unabhängig von der Entscheidung, welche Trainingssignaturen (»gute« oder »alle«) man aus dem Trainingsdatensatz auswählen sollte,

denn vorgegeben werden »optimale« Modelle, die durch Binärbildvorgabe bzw. der daraus berechneten Gradientenrichtungshistogramme erzeugten Daten definiert sind. Obendrein erspart man sich die fehleranfällige Maximasuche und bewahrt die weitgehende Kontrastunabhängigkeit.

Will man also z.B. solche Strukturen im Bild entdecken, die sich durch Parallelität *und* Orthogonalität (also Rechtecke) auszeichnen, so ist ein Modell wie in Abbildung 22 optimal.

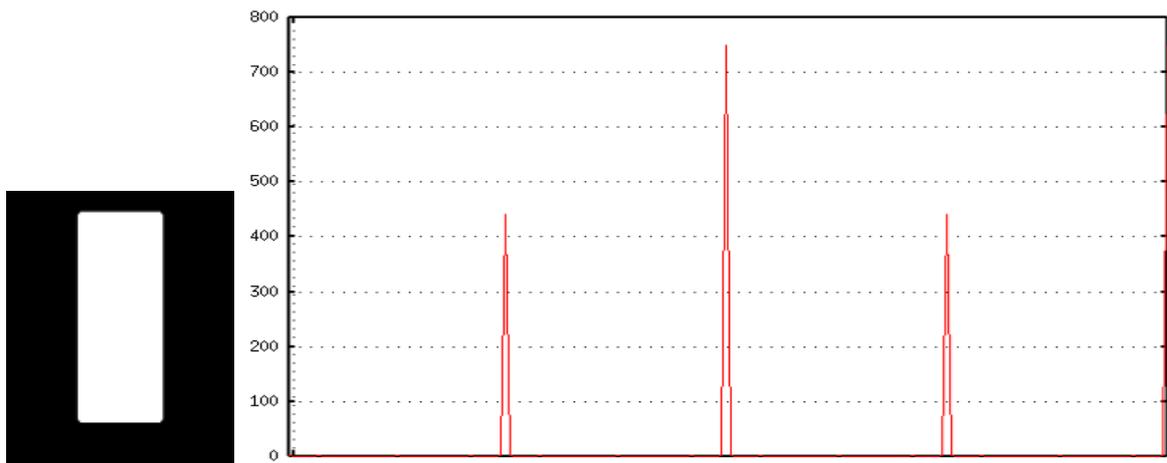


Abbildung 22: Optimales Rechteckmodell: weiß auf schwarzem, homogenem Grund (links), zugehöriges Gradientenrichtungshistogramm (rechts). Das vierte Maximum liegt auf der rechten Ordinate.

Hat man das gewünschte Modell zugrunde gelegt, korreliert man zuvor berechnete Gradientenrichtungshistogramme in verschiedenen Rotationsstufen (d.h. zyklischen Verschiebungen des Histogramms auf der Abszisse). Die Auflösung in Grad der Rotationsstufen ist anwendungsabhängig. Eine zwar gründliche, aber rechenintensive Analyse vergleicht in 1° -Schritten (d.h. 180 Histogramme) die Rechtecke des Modells mit den Gradientenrichtungshistogrammen lokaler Umgebungen bzw. Kacheln im Bild, die der Größe des gesuchten Zielobjekts entsprechen. Zur Beschleunigung der Vergleichsrechnung kann auf die FFT bzw. einen Korrelationskoeffizienten (25) zurückgegriffen werden, wobei auf die Normierung durch die Varianzen verzichtet werden kann.

$$c = \text{cov}(r_i, S), \quad (25)$$

mit c Vergleichsergebnis, r_i Gradientenrichtungshistogramm der jeweiligen lokalen Bildumgebung und S die Gradientenrichtungshistogramme (in verschiedenen Rotationen) des vorgegebenen Salienzmusters. Die Möglichkeit, mit Hilfe der Vergleichswerte (Korrelation/Kovarianzen) ein Maß für die Ähnlichkeit bzw. Auffälligkeit angeben zu können, ist auch für eine flexible Schwellwertfestlegung vorteilhaft. Die Prozedur zeigt sich im Überblick in Abbildung 23, wobei rote Pfeile/Kästchen »online«-Berechnungen und grüne Pfeile/Kästchen »offline«-Berechnungen markieren.

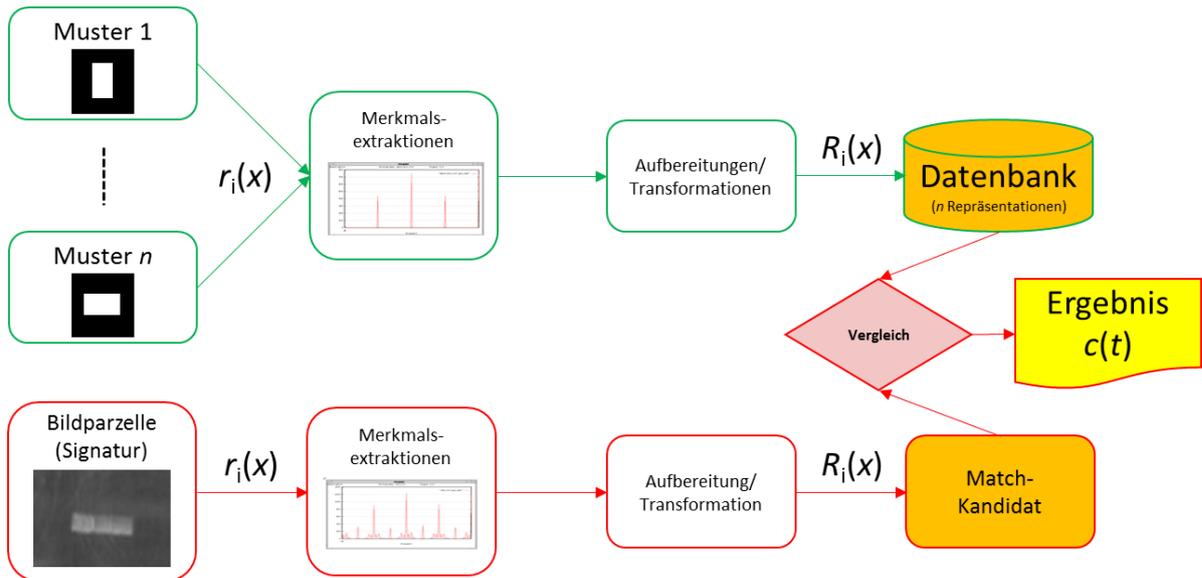


Abbildung 23: Vergleich der Gradientenrichtungshistogramme von lokalen Bildumgebungen und den Musterhistogrammen.

Die folgende Abbildung 24 zeigt ein Beispiel des Vergleichs eines Musterhistogramms mit einer Zielsignatur aus einer lokalen Bildumgebung.

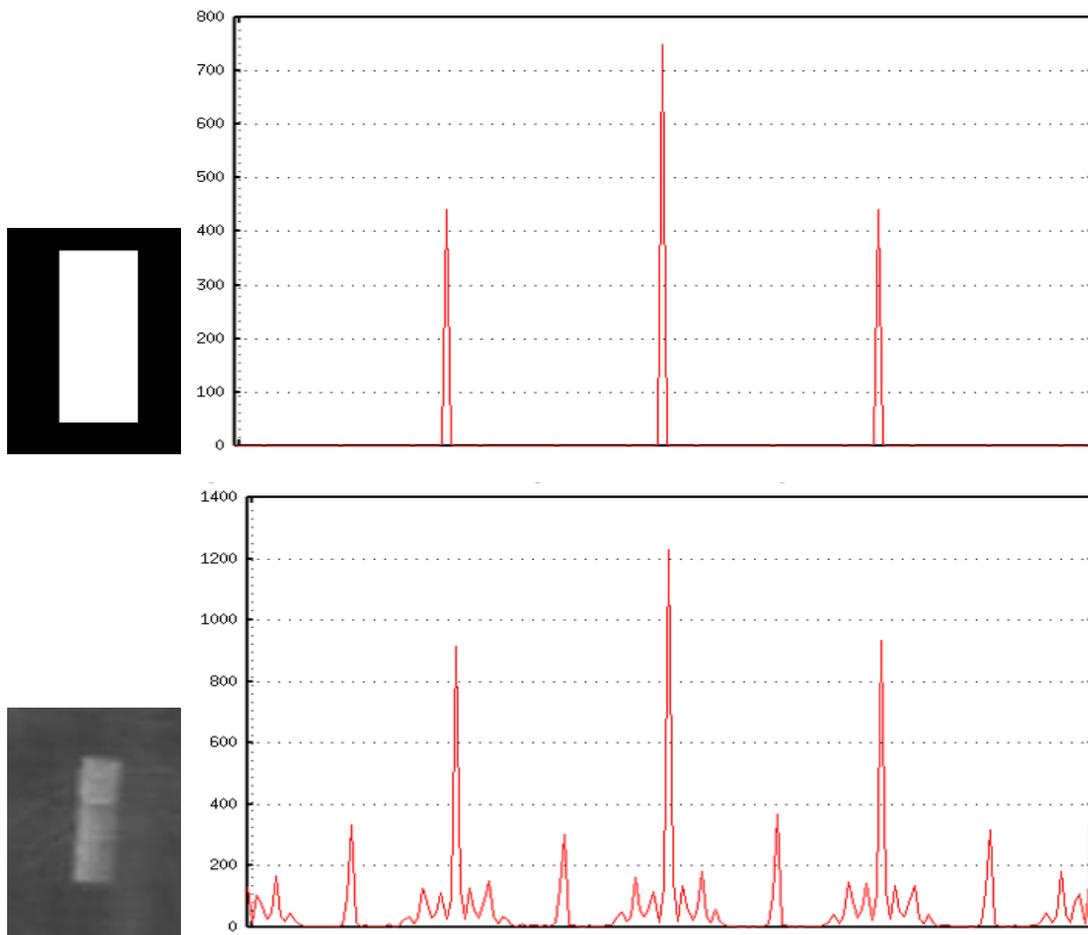


Abbildung 24: Vergleich – Muster und sein Histogramm (oben) sowie Zielsignatur in einem Bildbereich mit seinem Histogramm (unten).

Auch wenn das Histogramm des Bildbereichs verrauschter ist, so sind die Ähnlichkeiten auch visuell unmittelbar erkennbar.

Mit der Binärbildvorgabe und der entsprechenden Generierung der Gradientenrichtungshistogramme können Auffälligkeitsmaße, in Abhängigkeit der jeweiligen Aufgabenstellung bzw. Anwendung, nahezu beliebig definiert werden (vgl. Abbildung 25). Allein durch die Vorgabe entsprechender Binärmuster können auf einfache Art Salienzen definiert werden. Ferner können Binärmuster sogar zur Laufzeit eingespeist werden.

Der Ansatz über die Nutzung von Gradientenrichtungen von Binärmustern unterliegt jedoch auch Einschränkungen. Beispielsweise sind runde Strukturen oder Kreise mit ihren gleichverteilten Gradientenrichtungshistogrammen nur bedingt als zu detektierende Zielstruktur geeignet.

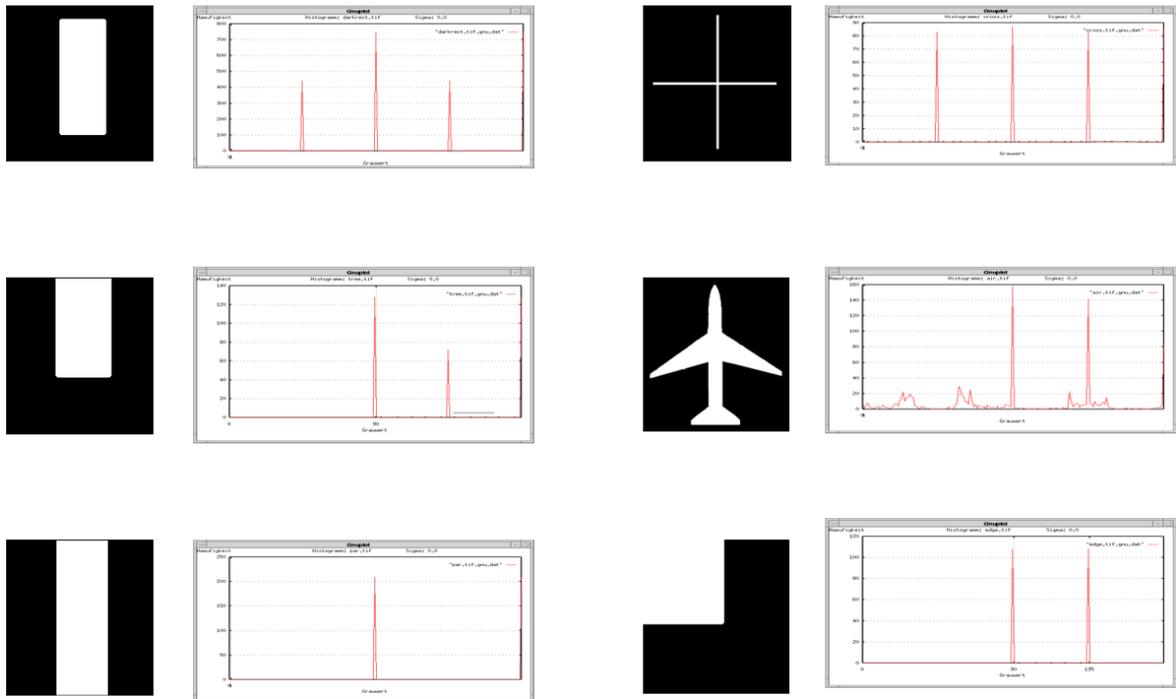


Abbildung 25: Übersicht einiger Beispilmuster mit ihren korrespondierenden Gradientenrichtungshistogrammen.

Dieser Ansatz ist sehr flexibel und mächtig, denn es müssen nur die grundlegenden Eigenschaften der gesuchten Primär- bzw. Sekundärartefakte (Salienzen) vorgegeben werden. Zeichnet sich die Salienz z.B. nur durch strukturelle Parallelität von Kanten aus, wie z.B. bei Straßen, Wegen und Spuren im Gelände oder Flußverläufen, so kann ein binäres Muster vorgegeben werden, welches im Gradientenrichtungshistogramm genau zwei Maxima im Abstand von 180° aufweist (Abbildung 26).

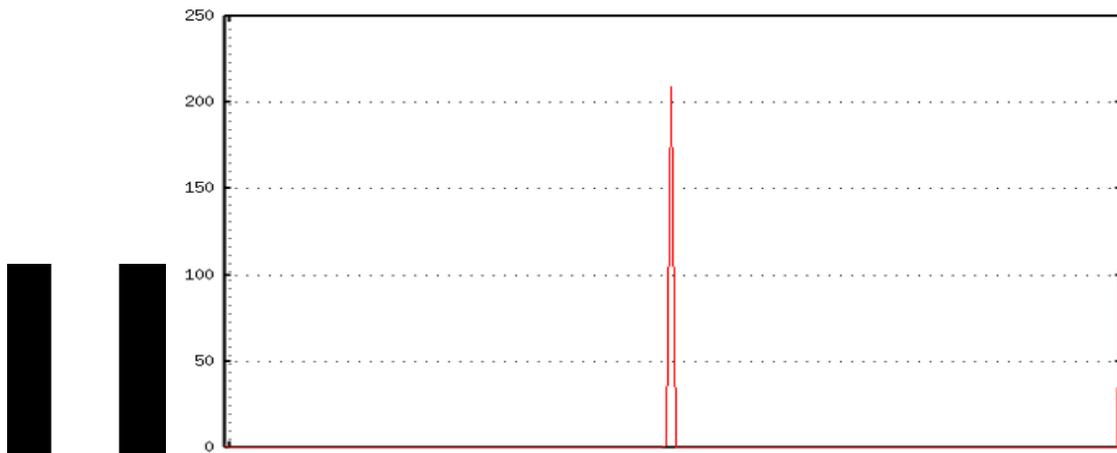


Abbildung 26: Muster zur Suche nach parallelen Strukturen (Straßen, Flüsse, ...)
Das Histogramm weist genau zwei Maxima im Abstand von 180° auf.

Die Suche nach parallelen Strukturen im Bild hat im Hinblick auf die Leistungsparameter der eigentlichen Detektionsaufgabe Vorteile. Beispielsweise befinden sich Fahrzeuge üblicherweise auf Straßen und Wegen oder hinterlassen im Gelände Spuren. Es bietet sich also an, in einer ersten Auswertestufe nach diesen Strukturen im Bild zu suchen und sie als »Erwartungsbereiche« (ROE – Region of Expectation) bzw. potentielle »Aufenthaltsbereiche« zu definieren. Eine ROE zeichnet sich also durch eine hohe Aufenthaltswahrscheinlichkeit von gesuchten Zielobjekten aus. Durch die Definition von ROEs liegen auch automatisch Ausschlussbereiche fest, die über eine geringe Aufenthaltswahrscheinlichkeit entsprechender Zielobjekte verfügen (Abbildung 27).

Detektiert man auf diese Weise z.B. Straßen, wird Kontextwissen generiert. Alternativ (oder ergänzend) können entsprechende Informationen a priori auch aus einem GIS abgerufen werden. Im Allgemeinen sind in einem GIS sogenannte »Straten«, also geographische Gebiete, die einem Homogenitätskriterium genügen, gespeichert. Je nach Anwendungsfall, können diese Straten zur Definition von ROEs herangezogen werden. Nachdem die ROEs festgelegt wurden, erfolgt im zweiten Schritt der eigentliche Detektionsvorgang.



Abbildung 27: Bild mit (aufgehellten) Erwartungsbereichen gemäß der Information »Straße«.

Die Suche nach Ecken bzw. eckförmigen Strukturen im Bild wird durch ein Richtungshistogramm ermöglicht, welches zwei Maxima im Abstand von 90° aufweist (Abbildung 28).

Bei der Verwendung der Histogramme wird nur auf die Ausprägung prominenter Maxima geachtet. Die absolute Höhe der Spikes ist nebensächlich. Solange sich die Maxima prominent ausprägen, spielen Rauscheffekte auch keine Rolle.

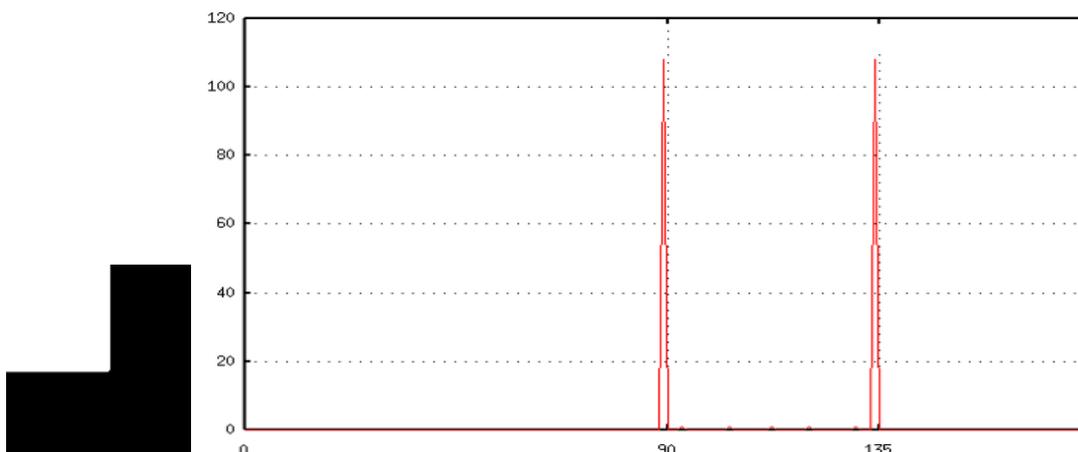


Abbildung 28: Beispielmustervorgabe zur Suche nach Ecken im Bildbereich.

Analog können auch komplexere Muster gesucht bzw. vorgegeben werden. Abbildung 29 zeigt die binarisierte Draufsicht eines Flugzeuges. Bei dem Richtungshistogramm bilden sich rauschartige »Huckel« aus. Diese entstehen an abgerundeten Ecken.

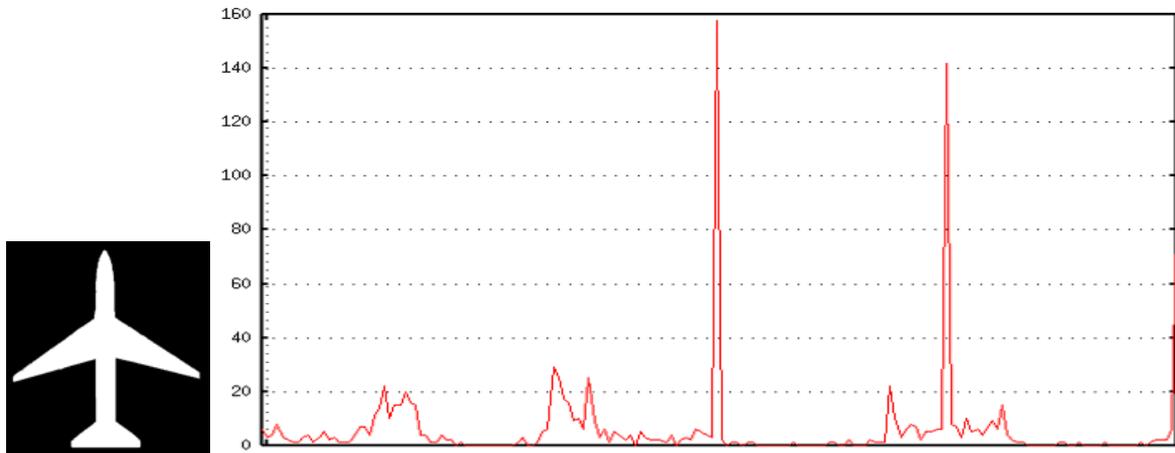


Abbildung 29: Vorgabemuster für ein Flugzeug.

Sei ein Bild gegeben und die Gradientenrichtungen berechnet. Die Aufgabenstellung erfordert die Suche nach Fahrzeugsignaturen in Senkrechtsicht. Gesucht wird in dem Fall also nach rechteckförmigen Zielsignaturen, womit das Rechteckmodell mit seinen 4 Maxima (2 lokale, 2 globale, im Abstand von 90°) zugrunde gelegt werden kann (vgl. Abbildung 22).

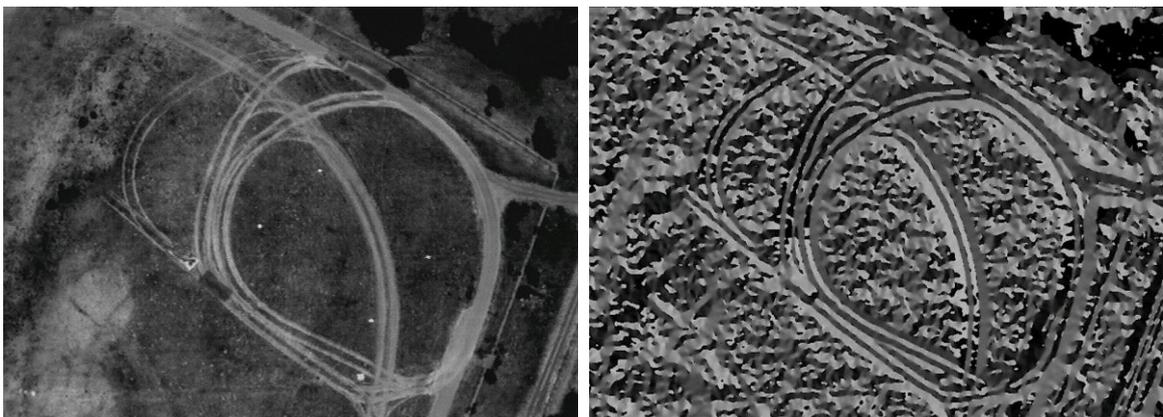


Abbildung 30: Beispielbild (links) nach Faltung und das grauwertcodierte Gradientenrichtungsbild, rechts.

Mit grafisch gezeichneten Vektoren kann man Gradientenrichtungen und –beträge gleichzeitig visualisieren. Beispielsweise, indem schachbrettmusterartig Punkte im

Bild gewählt werden und die grafischen Vektoren eingezeichnet werden (Abbildung 31). Die eingezeichnete Stärke des Vektors korrespondiert dabei mit der Größe des Gradientenbetrags. Mit dieser Visualisierung wird leicht erkannt, dass eher homogene Bildbereiche kleine grafische Vektoren aufweisen und Stellen mit hohem Kontrastunterschied große Vektoren.

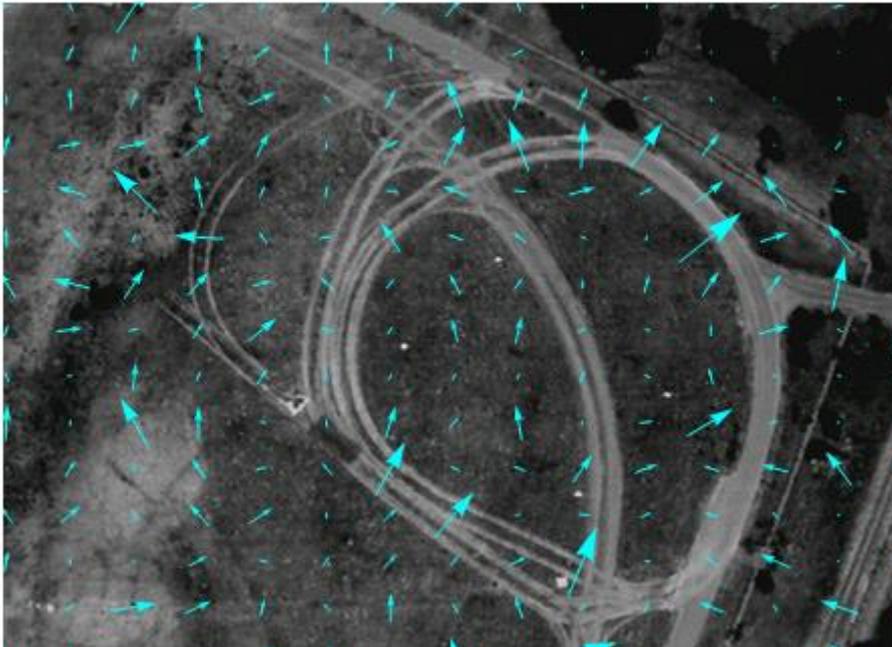


Abbildung 31: Gradientenrichtungen und -beträge gleichzeitig mittels grafischer Vektoren visualisiert.

Das Gradientenrichtungsbild kann – je nach zur Verfügung stehender Rechenleistung – dicht (im Rahmen dieser Arbeit der bevorzugte Ansatz) oder überlappend (gemäß ursprünglichem Korn'schen Ansatz) gekachelt bzw. parzelliert werden, wobei die Parzellen die Größe der gesuchten Zielsignatur haben. Für entsprechende Experimente ist im IOSB eine Anwendung (»Kacheln«) entwickelt worden, welche im Wesentlichen folgende Möglichkeiten bietet (vgl. Abbildung 32):

- Visualisierung des zugrunde gelegten Bildes oder des Gradientenrichtungsbildes.
- Interaktive Einstellung der Kachelgröße.
- Visualisierung der wählbaren Kachelung.
- Auflistung der Gütwerte (z.B. Korrelations-/Kovarianzwerte) für jede Kachel.
- Interaktive Einstellung eines Schwellwertes für die Gütwerte mit farblicher Visualisierung der Kacheln mit überschwelligem Gütwert.

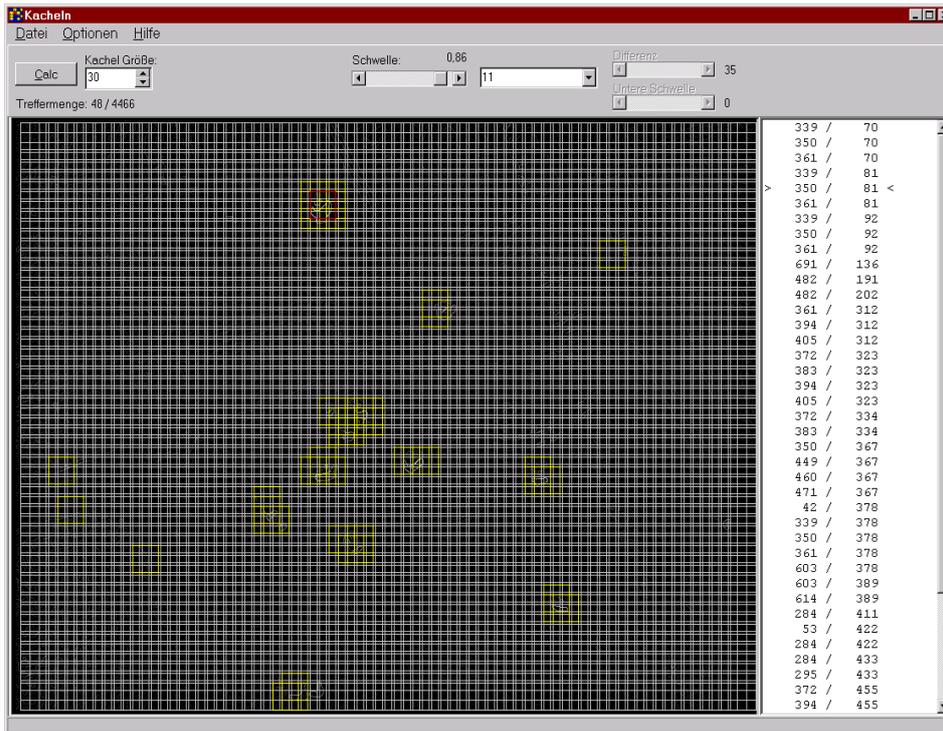


Abbildung 32: Anwendung »Kacheln« für Experimente bzgl. Schwellwerten von Gütemaßen, Kachelüberlappungen, etc. Kacheln mit einem überschwelligen Gütewert sind gelb gekennzeichnet. Bild: IOSB.

Als Ergebnisbeispiel legen wir Abbildung 33 inkl. Ground-Truth zugrunde. Der Korn'sche Ansatz operiert auf Bildern mit ausreichend ausgeprägtem Zielobjektkontrast gut, hat allerdings im Gegensatz zum hier vorgestellten Ansatz Probleme bei kontrastschwachen Objekten (Abbildung 35). Das trifft auch auf modernere Ansätze zu (z.B. SIFT), die im Falle kontrastschwacher Objekte versagen und daher keine Merkmale an der Zielsignatur im Bild identifizieren können (Abbildung 34). Weder beim SIFT-, noch beim Korn-Ansatz wird das obere Fahrzeug detektiert. Allerdings detektiert unser Ansatz auch Sekundärartefakte (Abbildung 36).



Abbildung 33: Grundwahrheit.

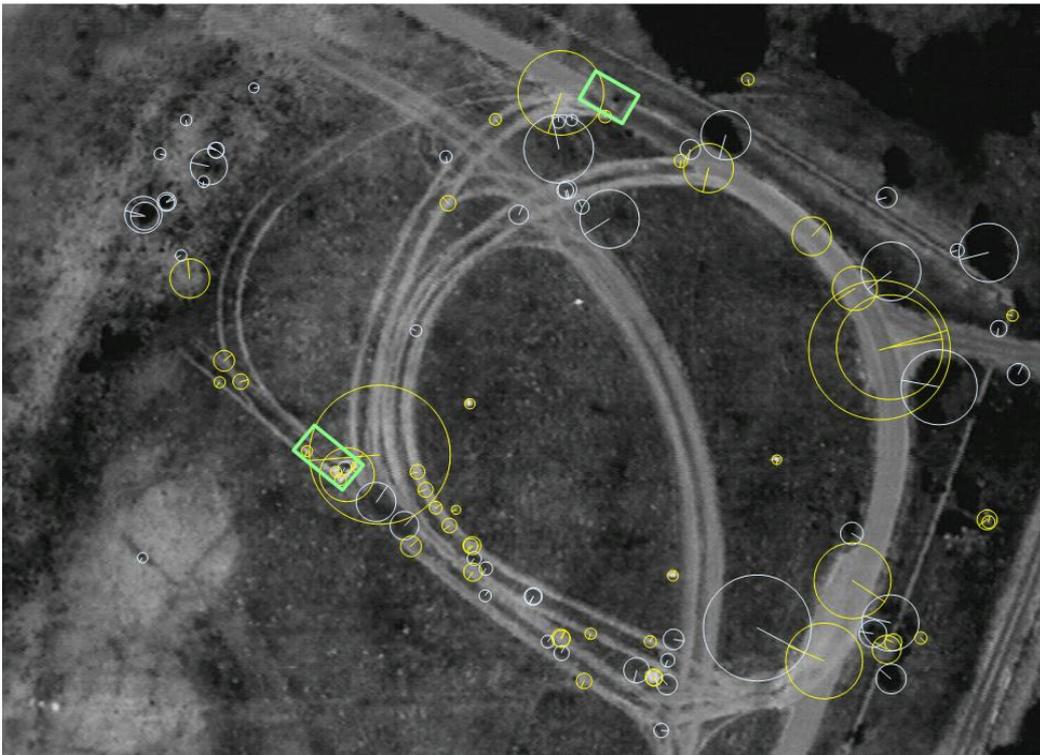


Abbildung 34: SIFT-Ansatz mit Ergebnissen und Grundwahrheit. Deutlich sichtbar ist, dass kein SIFT-Merkmal beim oberen Zielobjekt erfasst wird.



Abbildung 35: Ergebnis mit dem Korn-Verfahren.



Abbildung 36: Ergebnis mit eigenem Verfahren.

Die Eigenschaften des eigenen Ansatzes:

- Salienzen können durch einfache (Binär-) Modelle definiert werden.
- Sehr geringe Kontrastabhängigkeit.
- Durch die Vorausberechnung entsprechender Rotationen im 1D-Histogrammmodell wird eine Rotationsunabhängigkeit erreicht.
- Für jede Detektionshypothese wird ein Gütemaß generiert, was flexible Schwellwertanpassungen ermöglicht.
- Der Vergleich von Modell(histogrammen) und Bildbereich(histogrammen) erfolgt in 1D, was schneller ist als in 2D zu operieren.
- Eine fehleranfällige und aufwändige Maximasuche in Histogrammen entfällt.
- Der Prozess der Verfahrensoptimierung bzw. Parameteranpassungen wird durch die Nutzung der (Binär-) Modelle vereinfacht.

5 Beschleunigung der Verarbeitung

1965 formulierte GORDON E. MOORE das nach ihm benannte Moore'sche Gesetz (Moore, 1965), welches besagt, dass sich die Dichte Integrierter Schaltkreise (IC) alle 12 Monate verdoppelt. Abbildung 37 berücksichtigt eine Auswahl an CPUs der Firmen AMD, Intel, Motorola, MOS und Zilog zur Veranschaulichung dieser Entwicklung.

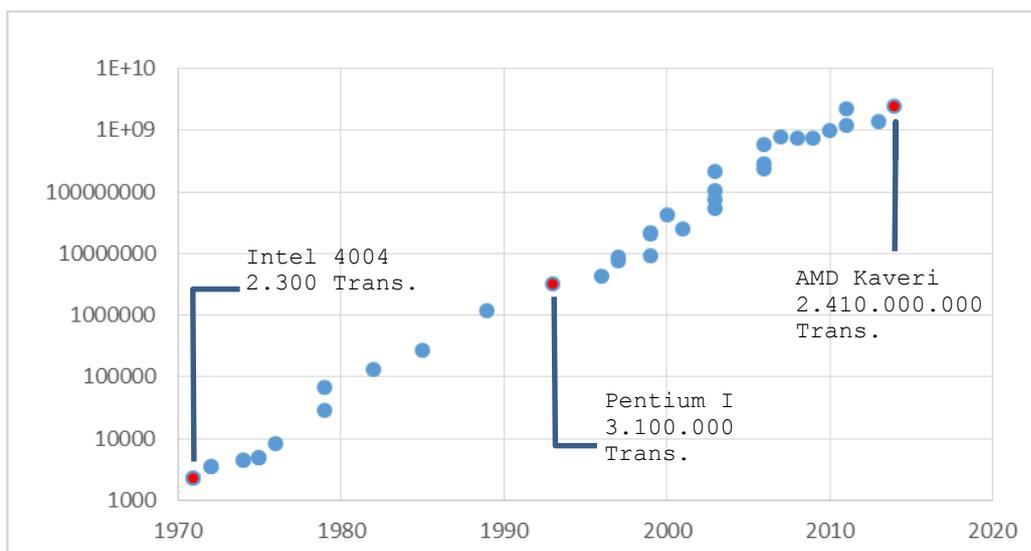


Abbildung 37: Anzahl an Transistoren in CPUs (logarithmische Skala).

Mit der Erhöhung der Integrationsdichte war gleichzeitig auch eine Steigerung der Taktfrequenz möglich. Dieses führte bis vor rund 10 Jahren dazu, dass insbesondere aufgrund höherer Taktraten leistungsgesteigerte Einzelprozessorsysteme gefertigt wurden. Viele Softwareentwickler überließen die Steigerung der Ausführungsgeschwindigkeit im Wesentlichen den hardwaretechnischen Fortschritten. Dieses Vorgehen konnte aus verschiedenen Gründen nicht funktionieren. So wies NIKLAUS WIRTH 1995 auf das nach ihm benannte Gesetz hin, dass »die Software schneller langsamer, als die Hardware schneller wird« (Wirth, 1995). Dieses Wirth'sche Gesetz warnt vor dem überproportionalen Anstieg der Komplexität von Programmen. Erschwerend kommt die exponentiell steigende Menge an zu verarbeitenden Daten hinzu. Es wird in diesem Zusammenhang von einem »Data Warehouse« (DWH) bzw. »Big Data« gesprochen (Oueslati, et al., 2010).

Im Video-/Multimedia-Bereich ist ein anhaltender Trend zu immer höheren Auflösungen zu verzeichnen. Operierte das Videocassettenformat VHS noch mit rund 0,08 MPixel Auflösung (pro Frame), sind in vielen Bereichen »Full HD«-Lösungen (ca. 2 MPixel/Frame) Standard. Erste Geräte mit »Quad Full HD« (auch 4K genannt) mit ca. 12,5 MPixel/Frame wurden bereits vorgestellt und Ankündigungen ranken sich um Systeme mit über 36 MPixel/Frame. Mit Blick auf die stark steigenden Datenmengen bzw. »Big Data« stellen wir in dieser Arbeit fest:

»Die zu verarbeitende Datenmenge wächst schneller, als Rechner schneller werden.«

Der wesentliche Grund, warum die Leistungssteigerung der CPUs primär aufgrund der Erhöhung der Taktrate von den Herstellern nicht in der bis vor rund 10 Jahren gewohnten Weise weiterverfolgt wurde, liegt am einhergehenden, steigenden Energieverbrauch der CPUs mit höherem Takt. Der Energieverbrauch in Abhängigkeit von der Taktfrequenz für eine Ein-Kern-CPU lässt sich wie folgt berechnen (Chandrakasan, et al., 1995). Sei C die Kapazität eines Schaltkreises, um Energie zu speichern, dann gilt $C = q/V$, mit q Ladung und V Spannung. Also gilt $q = CV$. Um eine Ladung q von 0 auf V zu bringen, wird die Arbeit $W=q \cdot V$ verrichtet. Es gilt also $W=CV^2$. Bei der Leistung wird vorliegend danach gefragt, wie oft man einen Schaltkreis pro Sekunde schalten lässt (Taktfrequenz), es gilt $P=f \cdot W$, mit f Taktfrequenz. Zusammengefasst ergibt sich also:

$$P = CV^2f \tag{26}$$

Wir nehmen für eine Ein-Kern-CPU die Werte als gegeben an und betrachten die Situation für eine Zwei-Kern-CPU mit einer nominell gleichen Rechenleistung für parallelisierbare Anwendungen. Die Gesamtkapazität der Zwei-Kern-CPU wird aufgrund des höheren Leitungsaufwandes etwas höher sein als 2, also z.B. 2,1. Die Taktfrequenz jedes Kerns kann halbiert werden, woraufhin auch die Spannung reduziert werden kann. Allerdings aufgrund von Leckströmen u.a. nicht ganz auf die Hälfte, sondern z.B. auf 0,6. Werden diese relativen Werte für Formel (26) zugrunde gelegt, errechnet sich ein Wert für P von ca. 0,38. Das bedeutet, dass das Zwei-Kern-Ensemble bei gleichem Durchsatz fast 40 Prozent weniger Stromverbrauch aufweist. Entsprechend sind die Werte bei CPUs mit bedeutend mehr Kernen (die »Xeon 5« CPU von INTEL verfügt bereits über 60 Kerne).

(Sutter, 2005) führte in seinem unter Softwareentwicklern viel beachteten Artikel »The Free Lunch Is Over« aus, dass seriell ausgelegte Hardware an ihre physikalischen Grenzen stößt und daher zwei wesentliche Konsequenzen zu ziehen seien:

1. Die CPU-Hersteller werden sich auf die Entwicklung von Multi-Kern-Prozessoren konzentrieren.
2. Software-Hersteller sind gezwungen, massiv parallele Programme zu entwickeln, um die Leistung von Multi-Kern-CPU's besser ausnutzen zu können.

Andererseits muss festgestellt werden, dass eine automatische Parallelisierung von Programmen bzw. Programmteilen trotz intensiver Forschungsbemühungen nach wie vor nicht funktioniert und möglicherweise auch nie funktionieren wird. Es liegt also am Programmentwickler, die Teile eines Programms zu identifizieren, die sich parallelisieren lassen. Dabei kann man im Wesentlichen auf eine Parallelisierung auf Prozessebene oder auf eine Vektorisierung setzen. Folgende Ansätze bieten sich an:

1. Spezielle Hardware, wie z.B. ASICs (Application-Specific Integrated Circuits), die sich jedoch erst ab höheren Stückzahlen (≈ 20.000) rechnen oder FPGAs (Field Programmable Gate Arrays). Beide Lösungen bieten eine sehr hohe Rechenleistung, die insbesondere bei vektorisierbaren Berechnungen die anderen Ansätze übertrifft. In beiden Fällen verlässt man jedoch den Bereich von Standard-PC-Hardware, was in vielen Anwendungsfällen nicht erwünscht ist.
2. Verarbeitung mittels GPU (Graphics Processing Unit), was jedoch entsprechende Grafikkarten erforderlich macht sowie nur bedingt portable Software (z.B. mittels OpenCL – Open Computing Language) oder gar proprietäre Software (z.B. mittels CUDA – Compute Unified Device Architecture von NVIDIA) bedeutet.
3. Parallelisierung auf Prozessebene, z.B. unter Nutzung von OpenMP (Open Multi-Processing), was in der vorliegenden Arbeit jedoch nur angedeutet wird (Abschnitt 5.2.1).
4. Nutzung spezieller Möglichkeiten heutiger Standard-CPU's, was im Rahmen dieser Arbeit ausgewählt wurde und im Abschnitt 5.2 besprochen wird.

5.1.1 Rechnerarchitekturen

Die hier gewählte Klassifizierung von Rechnerarchitekturen wurde erstmals von FLYNN vorgeschlagen (Flynn, 1972) und unterlag in der Folge diversen Verfeinerungen (Hellmann, 2013):

- SISD-Architektur (Single Instruction, Single Data): Diese Klasse von Rechnern realisiert das klassische sequentielle »John von Neumann«-Prinzip. Die auszuführenden Befehle (Befehlsstrom) werden einzeln nacheinander auf nacheinander zu beziehende Daten angewendet.
- MISD-Architektur (Multiple Instruction, Single Data): Rechner dieser Klasse bestünden aus einer Vielzahl von Prozessoren. Alle Prozessoren führen gleichzeitig verschiedene Befehle auf den Daten eines einzelnen Eingabestroms aus. Die Organisationskomplexität ist hoch und gleichzeitig besitzt diese Architektur durch die SD (Single Data)-Einschränkung nur eine geringe Flexibilität. Daher findet sich in der Praxis kein entsprechend organisiertes Rechnersystem.
- SIMD-Architektur (Single Instruction, Multiple Data): Bei reinen Rechnersystemen dieser Klasse findet man mehrere Prozessoren mit eigenem Speicher im System, die jeweils denselben Befehl auf unterschiedlichen Datenströmen ausführen. Diese Architektur ist für DSPs (Digital Signal Processor), systolische Arrays, Pipeline-Rechner und im übertragenen Sinne auch für SSE grundlegend. Für die Bild- und Signalverarbeitung besitzt sie eine besondere Bedeutung.
- MIMD-Architektur (Multiple Instruction, Multiple Data): Dieses ist das allgemeinste Architekturmodell. Das Rechnersystem besitzt mehrere Prozessoren, die unabhängig voneinander unterschiedliche (oder die gleichen) Daten verarbeiten können. Im Prinzip handelt es sich um eine Verschaltung von SISD-Rechnereinheiten. Bei geeigneter Konfiguration kann man mittels MIMD jede der genannten sonstigen Architekturen realisieren. In der Praxis findet man diese MIMD-Architektur mit gewisser Einschränkung sehr häufig schon bei Standardrechnern. Heutige Betriebssysteme unterstützen diese Architektur üblicherweise durch präemptives Multitasking.

Ein grundlegendes Problem bei allen Architekturen, in denen Parallelität eine Bedeutung, spielt liegt in der Frage nach der Parallelisierbarkeit von Berechnungen und insbesondere bei MIMD-Architekturen die Frage nach der Synchronisation von Vorgängen.

Die heute verfügbaren Mikroprozessoren realisieren zumeist mehrere Architekturen gleichzeitig. Grundsätzlich sind sie auch SISD-konform, verfügen jedoch über mehrere autarke »Kerne«, also separate CPUs auf einem Chip (MIMD-Prinzip) sowie eine SIMD-Erweiterung pro Kern (meist SSE).

5.2 SIMD-Realisierung der NAG

Aus dem weiten Feld der Forschung rund um die Parallelisierung von Algorithmen empfehlen sich bei vektorisierbaren Anwendungen insbesondere SIMD-Techniken. Unter SIMD versteht man im Allgemeinen eine bestimmte Klasse möglicher Rechnerarchitekturen. Die Firma INTEL hat ab 1997 mit MMX (Multi Media Extension) und ab 1999 SSE (Streaming SIMD Extensions) eingeführt und seitdem bis heute weiterentwickelt. Diese Verfügbarkeit in allen heutigen Standard-CPU's ist der Grund, warum die vorliegende Arbeit die Nutzung der SIMD-Erweiterungen der CPU's favorisiert.

Grob gesprochen kann bei einer SIMD-Architektur *ein* Befehl gleichzeitig auf n Daten angewendet werden. In der SSE-Realisierung können beispielsweise die Grundrechenarten oder auch das Skalarprodukt aus zwei – je nach Datentyp – aufgeteilten 256-bit Registern parallel berechnet werden. Der Vorteil für die Bildverarbeitung bei dieser Parallelisierung liegt auf der Hand: im Extremfall beherbergt ein 256-bit-Register die Werte von 256 Pixeln eines Binärbildes, die dann mit einer logischen Operation parallel mit anderen 256 Pixeln verrechnet werden können. In diesem speziellen Fall sind Berechnungsgeschwindigkeiten erreichbar, die bei gleichem CPU-Takt um 1-2 Größenordnungen höher liegen als bei der sequentiellen Abarbeitung. SIMD-Architekturen sind für die Bild- und Signalverarbeitung von besonderer Bedeutung, da bei sehr vielen Operationen die Werte der Bildmatrix parallel verarbeitet werden

können (z.B. bei einer Faltung oder generell bei lokalen Bildoperatoren). Diese inhärente Parallelität kann jedoch, trotz intensiver Forschung auf diesem Gebiet, nicht von gängigen Compiler-Systemen automatisch erkannt und umgesetzt werden. Daher muss die Spezifikation und Implementierung dediziert bzw. unter Nutzung entsprechender Entwicklungsumgebungen erfolgen.

5.2.1 SIMD-Realisierung des veränderten NAG-Operators

Der NAG-Operator basiert im Wesentlichen auf zwei Faltungen und den Berechnungen der Gradientenwinkel sowie Gradientenbeträge aus den Faltungsergebnissen. Aus Effizienzgründen werden die Winkel in 2-Grad-Schritten codiert (um in einem Byte erfassbar zu sein). Ferner wird aus Effizienzgründen bei der Berechnung der Winkel und Beträge auf eine Look-up-Tabelle mit entsprechend vorausberechneten Werten zurückgegriffen. Ebenso wurde auf ganzzahlige Faltungsmasken mit optimierten Werten zurückgegriffen. Diese Faltungsmasken sind 5×5 groß und lehnen sich daher an $\sigma \approx 0,9$ an, vgl. Formel (21). Dieser Wert von σ wurde gewählt, weil damit eine effiziente (separable), ganzzahlige Implementierung ermöglicht wird und sich eine solche Maskengröße für viele praktische Anwendungen K^T als geeignet erwies. Andere Werte für σ lassen sich über entsprechende Skalierungen des Eingangsbildes realisieren. Die Werte der Faltungskerne K bzw. K^T lauten wie folgt:

$$K = \begin{pmatrix} -1 & -3 & 0 & 3 & 1 \\ -8 & -24 & 0 & 24 & 8 \\ -14 & -42 & 0 & 42 & 14 \\ -8 & -24 & 0 & 24 & 8 \\ -1 & -3 & 0 & 3 & 1 \end{pmatrix}, K^T = \begin{pmatrix} -1 & -8 & -14 & -8 & -1 \\ -3 & -24 & -42 & -24 & -3 \\ 0 & 0 & 0 & 0 & 0 \\ 3 & 24 & 42 & 24 & 3 \\ 1 & 8 & 14 & 8 & 1 \end{pmatrix}$$

Nach den zwei Faltungen müssen noch die Gradientenbeträge, deren Maxima und die Gradientenrichtungen berechnet werden. Außerdem sind Koordinatentransformationen (kartesisch \rightarrow polar) vorzunehmen. Zur Vermeidung der Berechnung von transzendenten Funktionen wurden umfangreiche Funktionswerttabellen verwendet, die durch ein eigenes Programm generiert und fest codiert wurden. Aufgrund der Separierbarkeit und mehrfach vorkommender Koeffizienten konnten diverse Vereinfachungen vorgenommen werden. Zusammen mit Nebenrechnungen (Umwandlungsroutinen usw.) werden ca. 100 Grundoperationen pro Bildpunkt ausgeführt.

Aus Kompatibilitätsgründen wurde nur auf den MMX-Standard zurückgegriffen. Die Nutzung von SSE-Instruktionen und -Möglichkeiten erbrachte keine signifikante Geschwindigkeitssteigerung (Ausnahme: Maxima-Filterung, siehe unten). Die Kompatibilität mit dem MMX-Standard ermöglicht grundsätzlich eine Nutzung der SIMD-NAG-Realisierung im Embedded-Bereich.

Der Geschwindigkeitsgewinn durch die SIMD-Implementierung im Vergleich zur SISD-Implementierung liegt bei rund 300% (Abbildung 38). Allerdings verbraucht die vorliegende Implementierung noch gut 50% der Rechenzeit für die Herausfilterung der Maxima aus den Faltungsergebnissen. Die Filterung der Maxima könnte mit entsprechenden Instruktionen ab der Version 4.1 von SSE (INTEL, 2007) weiter beschleunigt werden, so dass ein Geschwindigkeitsgewinn von geschätzten 400-500% realisierbar wäre.

	Pentium I MMX 200 MHz	INTEL i7 3,45 GHz
SISD	0,5 Mio. <u>Pix/s</u>	35 Mio. <u>Pix/s</u>
SIMD	1,6 Mio. <u>Pix/s</u>	100 Mio. <u>Pix/s</u>

Abbildung 38: Leistungsvergleich SISD und SIMD.

Bemerkungen:

- Im Vergleich zur ersten MMX-CPU von 1997 erzielen heutige CPUs eine um etwa den Faktor 60 höhere Durchsatzsteigerung. Demgegenüber hat sich der CPU-Takt nur um etwa den Faktor 17 gesteigert. Der im Vergleich zur Taktsteigerung überproportionale Gewinn resultiert im Wesentlichen aus vergrößerten Registern und optimierten Schaltungslayouts.
- Die SIMD-NAG rechnet auf einem Kern. Auch der Leistungsvergleich wurde nur auf einem Kern durchgeführt. Daher ist die NAG-Anwendung nahezu linear skalierbar, eine Zwei-Kern-CPU liefert nahezu die doppelte Leistung.
- Mit der SIMD-NAG ist es heute möglich, Leistungswerte zu erzielen, die erst CPUs der nächsten oder übernächsten Generation ohne die SIMD-Implementierung erreichen.

6 Videogestützte Sicherheitstechnik

Die umfassende Thematik »Sicherheit« gehört zu den zentralen gesellschaftlichen Themen, mit denen sich politische bzw. behördliche Entscheidungsträger ebenso befassen, wie Vertreter aus der Wirtschaft und der Wissenschaft. Darüber hinaus sieht sich auch der einzelne Bürger mit diesem Themenkreis konfrontiert, sei es im Hinblick auf den Schutz vor Kriminalität, Naturgefahren, Terrorismus, Havarien, Epidemien, bis hin zur Wahrung seiner Persönlichkeitsrechte in einer informationstechnologisch vernetzten Welt. Bei der Realisierung von Sicherheitssystemen sind nationale und internationale gesetzliche Rahmenbedingungen sowie der Datenschutz ebenso zu beachten wie Aspekte der Normung (z.B. DIN, EU), der Interoperabilität, systemtechnische Aspekte, Zuverlässigkeit und der Ergonomie/Bedienbarkeit. Natürlich gibt es viele sicherheitstechnologisch relevante Forschungsgebiete; der vorliegende Schwerpunkt in Sachen »Sicherheitstechnik« basiert jedoch auf Bildfolgen.

6.1 VSAM – Video Surveillance and Monitoring

6.1.1 Genereller Aufbau

Eines der ersten größer angelegten Projekte zum Thema intelligenter Videosicherheitstechnik war das von 1997-1999 laufende VSAM⁸-Vorhaben (Collins, et al., 2000). In VSAM wurde ein Netzwerk von intelligenten Kameras zur kontinuierlichen und großflächigen Erfassung von Personen und Fahrzeugen in dynamischer Umgebung zugrunde gelegt, um ein durch eine Einzelperson bedienbares Sicherheitssystem zu konstruieren. Das Projekt endete vorläufig mit einem Experimentalsystem, das nahezu alle bis heute für Sicherheitssysteme relevanten Aspekte ansprach. Aus diesem Grund und auch weil es bis heute weiterentwickelt wird, ist es als eine Art Referenzsystem geeignet (Lakshmi Devasena, et al., 2011).

⁸ VSAM: Video Surveillance and Monitoring

Das VSAM-Experimentalsystem (Testbed) bestand aus einer zentralen Kontrolleinheit (OCU – Operator Control Unit), die u.a. Sensor- und Kollateraldaten empfing und verarbeitete, Datenmodelle bereit hielt, für die weitere Verarbeitung sorgte und für die Ergebnisvisualisierung sowie Interaktion eine Benutzungsschnittstelle aufwies. Damit sollten Bildfolgenauswerteverfahren für diverse Anwendungsszenarien geeignet demonstriert werden. Ein weiterer wichtiger Aspekt der aufgabengesteuerten Kontrolleinheit bestand in der Sensoreinsatzplanung. Für die nötigen Sensorströme sorgte ein Videokameranetz auf dem Campus der Carnegie Mellon University.

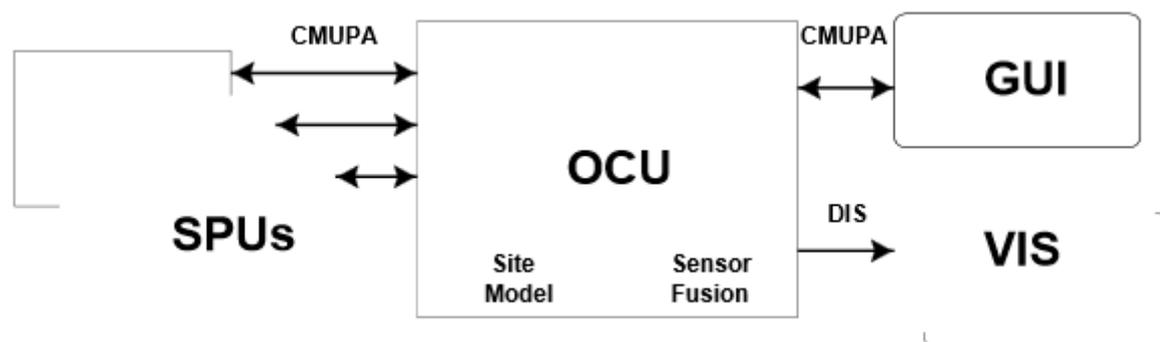


Abbildung 39: Architekturübersicht des VSAM-Systems (Collins, et al., 2000)⁹.

An die zentrale Kontrolleinheit (OCU) waren intelligente Kameras (SPU – Sensor Processing Units) angeschlossen, deren eingebaute Verarbeitungseinheit als Filter agieren sollte. Die Aufgabe bestand in der Detektion relevanter Ereignisse und der Übermittlung symbolischer Daten an die Kontrolleinheit. Dadurch sollte insbesondere die Netzwerklast so weit wie möglich reduziert werden. Die einzelnen SPUs stellten ein heterogenes Netz unterschiedlicher Kameratypen dar (u.a. thermisches Infrarot und visuell-optische Systeme). Zusätzlich zu den festinstallierten SPUs nutzte VSAM auch eine fahrzeug- sowie eine luftgestützte SPU.

Die aufgabengesteuerte Sensoreinsatzplanung der OCU basierte auf einer Kostenfunktion. Diese Kostenfunktion bewertete die detektierten Ereignisse sowie die Wer-

⁹ CMUPA: Carnegie Mellon University Packet Architecture, DIS – Distributed Interactive Simulation (Kommunikationsprotokolle), VIS (hier): Visualisierung.

tigkeit der Aufgabenstellung und berechnete die Kosten des jeweiligen Sensoreinsatzes. Auch für diese Abschätzungen war ein digitales Elevationsmodell der zu überwachenden Szene notwendig.

Diskussion:

Die Ziele von VSAM, die wesentliche Verarbeitung schon in den angeschlossenen intelligenten Kameras durchzuführen, um insbesondere relevante Ereignisse als solche automatisch zu erkennen und nur noch symbolische Ergebnisdaten an eine Zentrale zu übermitteln sind zwar verständlich, aber auch heute noch, bis auf spezielle Anwendungen, z.B. im Fahrerassistenzbereich, zu ambitioniert. Im Grunde werden die generellen Herausforderungen an die Bildauswertung als »gelöst« betrachtet und in die intelligenten Kameras verlagert. Ein Ansatz, der die Machbarkeiten in Teilen verkannte und der dafür sorgte, dass im Rahmen von VSAM viel Entwicklungsarbeit in hochperformante, allerdings jedoch dedizierte und unflexible Hardware investiert wurde.

Die aufgabengestützte Abarbeitung und Kostenschätzung hat in VSAM ebenso dedizierten Charakter. Es konnten keine Auswerte- bzw. Überwachungsaufgaben (z.B. bestimmtes Fahrzeug oder bestimmte Person verfolgen) formuliert werden, sondern mussten durch Anpassungen der Verarbeitungsketten von Hand umgesetzt werden.

7 Videoauswertesystem ABUL

Das vom Autor erdachte und im Fraunhofer IOSB entwickelte Videoauswertesystem ABUL¹⁰ ist für Sicherheitsanwendungen unter Nutzung der bildgebenden Sensorik unbemannter Luftfahrzeuge konzipiert und entwickelt worden (Müller, et al., 2013). Ursprünglich wurde es konzipiert, um Bildfolgenverarbeitungs- und -auswerteverfahren in einem System- und Anwendungszusammenhang testen, demonstrieren und weiterentwickeln zu können. Damit knüpft es an frühere Arbeiten der 80er und 90er Jahre im IOSB (vormals IITB) an, die unter dem Titel KIBASX¹¹ erfolgten und, im Gegensatz zu ABUL, verstärkt Bildauswertung auf Einzelbildern statt Videobildfolgen unter den Aspekten der Mensch-Maschine-Interaktion, Ergonomie und Vernetzung zum Ziel hatten (Schönbein, 1989). Beiden Systemen gemeinsam ist der generische Ansatz, welcher die leichte Integrierbarkeit neuer Auswerteverfahren ermöglichen soll. Dazu gibt es Konfigurationsmodelle, nach außen offen gelegte Software-schnittstellen sowie die Unterstützung einer breiten Palette von Standardformaten. Damit wird es auch Dritten ermöglicht, eigene Algorithmen in das Auswertesystem zu integrieren. Für die Bewertung von Tarnmaßnahmen, also Ansätzen zur Unkenntlichmachung von Objektsignaturen, wurde ABUL um Funktionen zur Tarnbewertung erweitert und CART¹² genannt (Müller, et al., 2011). Ferner erfuhr ABUL eine Erweiterung durch eine Videodatenbank, welche VABUL¹³ genannt wird (Brüstle, et al., 2012).

¹⁰ ABUL: Automatische Bildauswertung für unbemannte Luftfahrzeuge

¹¹ KIBASX: Konfigurierbares, interaktives Bildauswertesystem, basierend auf X-Window

¹² CART: Camouflage Assessment in Real Time

¹³ VABUL: Videodatenbank ABUL

7.1 Architektur

Die Kommunikationsarchitektur von ABUL ist in 3 Hauptkomponenten und ein übergeordnetes Steuermodul aufgeteilt. Technisch erfolgt die Kommunikation über einen Dbus (Desktop-Bus), welcher als Vermittlungssoftware (engl. *middleware*) einen freien IPC¹⁴-Rahmen realisiert (Esswein, 2012). Insbesondere ist damit auch ein ORB¹⁵ umgesetzt, womit die plattformunabhängige Kommunikation verteilter und parallel ablaufender Anwendungen ermöglicht wird (Abbildung 40). Der Dbus stellt in Teilen eine freie Software nach dem CORBA¹⁶-Vorbild der OMG¹⁷ dar. Durch die Nutzung einer solchen Kommunikationsinfrastruktur wird die Integration neuer Komponenten in ABUL im verteilten, heterogenen Verbund deutlich vereinfacht.

¹⁴ IPC: Interprozesskommunikation

¹⁵ ORB: Object Request Broker

¹⁶ CORBA: Common Object Request Broker Architecture

¹⁷ OMG: Object Management Group

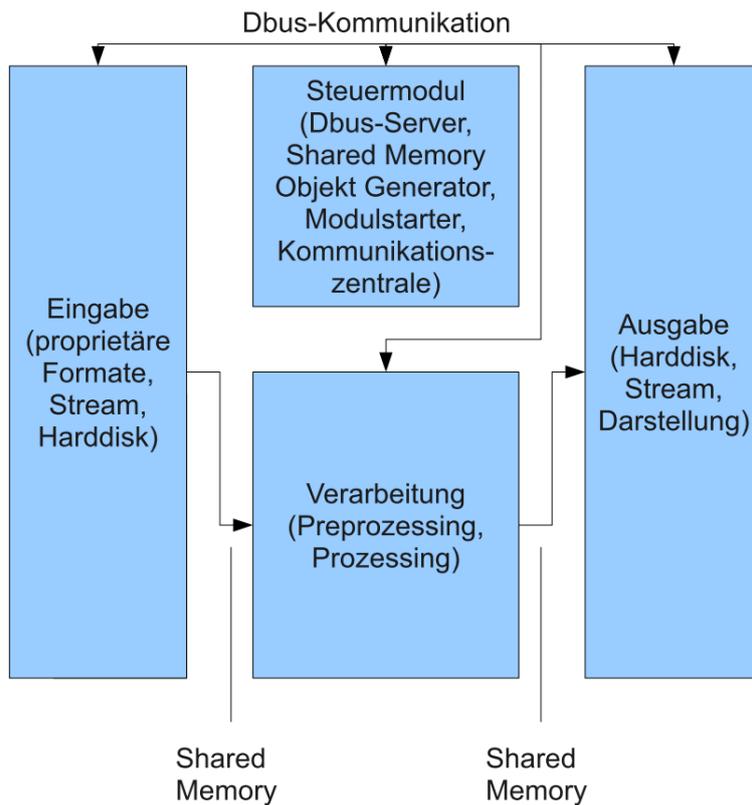


Abbildung 40: Aufbau der Kommunikationsstruktur in ABUL. Bild: IOSB.

7.2 Benutzungsschnittstelle

Die zwei grundlegenden Benutzungsschnittstellen von ABUL bestehen aus einem Video- und einem Auswertefenster. Im Videofenster (Abbildung 41) werden die verfügbaren Bilddatenströme (in Abhängigkeit der Sensorausstattung des UAV) in miniaturisierter Form angezeigt und zur Auswahl angeboten. Der jeweils selektierte Videostream erscheint in voller Auflösung mittig. Das Fenster bietet die generellen Funktionen moderner Videobetrachter und -rekorder, inkl. des zeitversetzten Betrachtens einer in Echtzeit einkommenden Bildfolge (engl. *time shift*) sowie gängige Bildmanipulationsroutinen (Kontrast, Helligkeit, etc.) Ferner werden hier die Bildverarbeitungsverfahren (z.B. Bildteppichgenerierung) ausgewählt.

Das Auswertefenster (Abbildung 42) visualisiert die Ergebnisse der Verfahrensauswahl und zeigt eine Zeitleiste, um in der erfassten Bildfolge zu navigieren, einzelne Abschnitte zu markieren oder auszuschneiden.

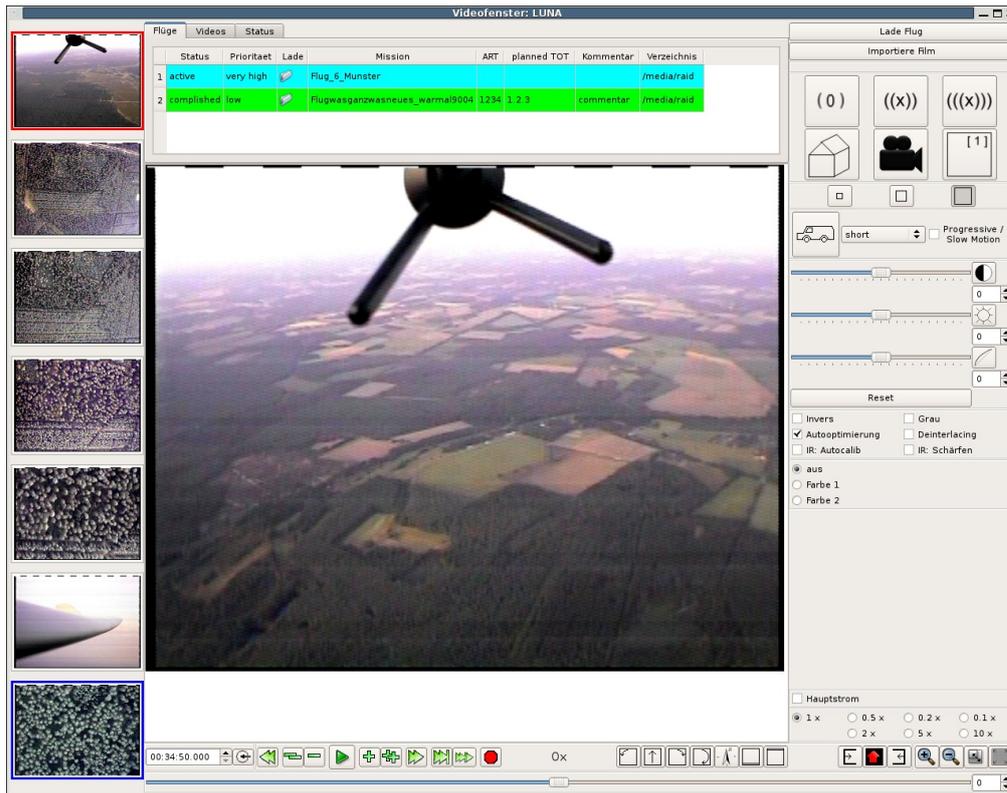


Abbildung 41: Benutzungsschnittstelle Videofenster. Bild: IOSB.

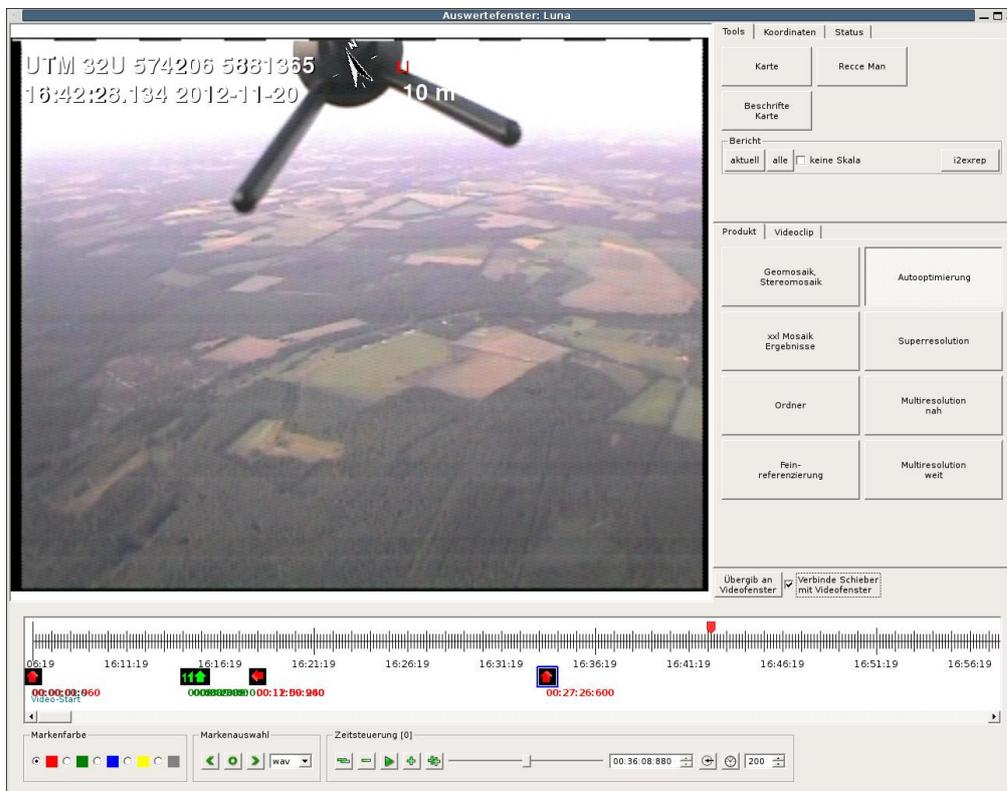


Abbildung 42: Benutzungsschnittstelle Auswertefenster. Bild: IOSB.

In der Echtzeitverarbeitung bleibt dem Operator wenig Zeit, um sich mit der Auswertung ggf. relevanter Szenenbereiche zu beschäftigen. Daher benötigt er eine Mimik, die ihm die Markierung interessierender Bildfolgenbereiche mit einer Start- und Endpunktmarke erlaubt. Die Ergebnisse werden zusammen mit der Bildfolge gespeichert, um für eine spätere oder, übergeben an andere Auswertinstanzen, parallel laufende Detailauswertung verwendet werden zu können (Abbildung 43). Die Mimik zum Setzen der Marken erlaubt ferner die Navigation im Bildstrom.



Abbildung 43: Definition von Marken im Bildfluss zur Definition von relevanten Teilabschnitten. Bild: IOSB.

7.3 Bildauswerteverfahren

Im Folgenden werden exemplarisch einige Bildverarbeitungs- und -auswerteverfahren vorgestellt. Im Zentrum stehen Registrierverfahren und ihre Anwendungen.

7.3.1 Bildregistrierung

Die Registrierung konsekutiver Einzelbilder (auch Bildstapelung oder Bildverrechnung genannt) einer Bildfolge ist von zentraler Bedeutung bei bewegten Sensorplattformen. Ein präzises Registrierungsergebnis macht viele interaktive oder automatische Auswertevorgänge überhaupt erst möglich. Beispielsweise kommt es bei mobilen, von Haus aus nicht stabilisierten Sensorplattformen häufig zu Wackeleffekten, die den menschlichen Beobachter stören und ermüden. Auch bei ortsfesten Sensoren kommt es häufig zu hochfrequenten Zitter- bis hin zu kapitaleren Wackeleffekten, z.B. bei Mastmontagen. Um zwei Bilder aufeinander zu registrieren, kommen grundsätzlich Verfahren in Frage, die

- 0-dimensionale (Punkte, Ecken)
- 1-dimensionale (Geradensegmente/Kurven)
- 2-dimensionale (Flächen, z.B. durch Polygonzüge definiert) oder
- 3- bzw. 2½-dimensionale (z.B. Elevationsmodelle)

Merkmale nutzen. Dabei werden die jeweiligen Merkmale aus den zwei zu registrierenden Bildern extrahiert, eine Korrespondenzanalyse durchgeführt und daraus dann die Parameter für eine Transformationsfunktion (z.B. affin, projektiv) geschätzt. Die höherdimensionalen Merkmale stellen höhere Anforderungen an ihre Extrahierbarkeit im Hinblick auf die Szenencharakteristik. Beispielsweise sind Geradensegmente als Straßenhypothesen bei großflächigen Aufnahmen z.B. in mitteleuropäischen, also bewohnten, infrastrukturell ausreichend ausgestatteten Szenen eine gute Wahl, da entsprechende Kanten- bzw. Liniendetektoren (Krüger, 2001) zumeist ausreichend viele und über das Bild verteilte Merkmale extrahieren können (Müller, et al., 1996). Bei anderen Szenen bzw. entsprechend kleinen Erfassungsbereichen des Sensors, wie es bei Videosensorik häufig der Fall ist, lassen sich oft nur abstraktere Merkmale (Punkte) finden (Müller, et al., 2001). Erste Arbeiten befassten sich mit der Anwendung des »Sternbildalgorithmus« (Zimmermann, et al., 1989) und dessen Anwendbarkeit auf die Bildregistrierung (Müller, et al., 1997). Bei der Ergebnisvisualisierung werden die Zuordnungen der korrespondierenden Bildmerkmale als Vektorfeld dargestellt und das Registrierungsergebnis in anaglyphischer Darstellung visualisiert (Abbildung 44).

Die Extraktion von punktförmigen Merkmalen zur Lösung der Registrierungsaufgabe stellt nach wie vor ein Forschungsthema dar. Dabei geht es um die Verbesserung der Robustheit sowie der Effizienz der Extraktion. Dem Sternbildalgorithmus liegen im Wesentlichen Filterungen mit einem Bandpass in 3 Stufen »fein«, »mittel« und »grob« zugrunde. Neuere wollen ein höheres Maß an Invarianz gegenüber einem ganzen Bündel an Einflussfaktoren erreichen. Beispiele dafür sind die SIFT¹⁸- und SURF¹⁹-Ansätze. Der SIFT-Ansatz, vorgeschlagen von (Lowe, 2004), konzentriert

¹⁸ SIFT: Scale Invariant Feature Transform

¹⁹ SURF: Speed-Up Robust Features

sich auf die Beherrschung der Rotation, Skalierung, affinen Deformationen, Aufnahmewinkeländerungen, Rauschen sowie Beleuchtungsänderungen. Das analoge Ziel verfolgt auch der SURF-Ansatz. Es stellt sich dabei heraus, dass der SIFT-Ansatz mehr robuste Merkmale aus dem Bildbereich extrahieren kann, dafür aber auch länger benötigt (Panchal, et al., 2013):

Algorithmus	Merkmale Bild1	Merkmale Bild2	Robust ²⁰	Rechenzeit
SIFT	892	934	41	1,543 s
SURF	281	245	28	0,546 s

²⁰ Damit wird angegeben, wie viele Merkmale tatsächlich korrekt zugeordnet wurden.

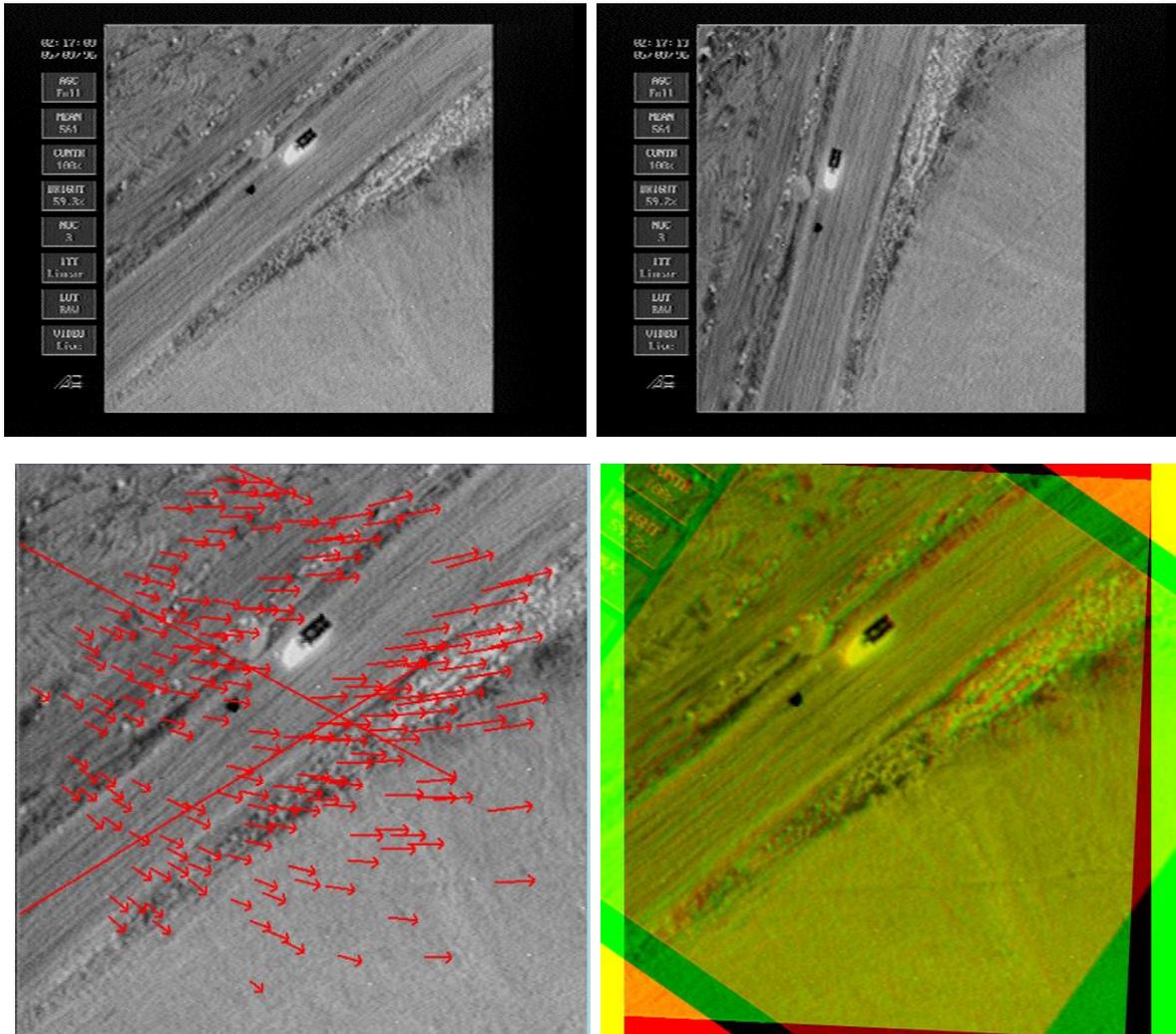


Abbildung 44: Zwei Beispielaufnahmen derselben Szene aus unterschiedlichen Winkeln (oben), Verschiebungsvektorfeld (unten links), Registrierungsresultat in anaglyphischer Darstellung (unten rechts). Bilder: IOSB.

Für die Berechnungen wurden ein Bildpaar mit annähernder VGA-Auflösung (640 x 478) und ein Rechner mit Intel i3-CPU (2,3 GHz) zugrunde gelegt. Abgesehen davon, dass der SIFT-Ansatz gegenüber SURF die dreifache Rechenzeit beansprucht, wird deutlich, dass beide Ansätze für die Videoauswertung in Echtzeit ungeeignet sind. Dafür sind die Rechenzeiten selbst des schnelleren SURF-Ansatzes mit über $\frac{1}{2}$ Sekunde für ein Bildpaar zu lang.

In das noch vorzustellende Videoauswertesystem ist ein Ansatz zur Merkmalsextraktion und Punktkorrespondenzschätzung integriert, der zwar nicht die Robustheit

wie SIFT/SURF aufweist (insbes. im Hinblick auf Aufnahmewinkelvariationen), dafür jedoch echtzeitfähig operiert (Krüger, 2001), (Müller, et al., 2007), (Heinze, et al., 2010). Die höhere Empfindlichkeit gegenüber Aufnahmewinkelveränderungen kann toleriert werden, da sich in den typischen Szenarien (z.B. Überflug) keine starken bzw. sprunghaften Winkeländerungen ergeben.

7.3.2 Bildteppiche

In vielen Fällen erfasst ein bildgebender Sensor nur einen kleinen Boden- bzw. Szenenbereich. Dabei geht rasch die Übersicht verloren und Objekte, die größer als der Erfassungsbereich des Sensors sind, können nicht als Ganzes aufgenommen werden. Das kann, in Abhängigkeit der Aufnahmeparameter, beispielsweise bei einem Überflug über ein größeres Schiff oder eine größere Liegenschaft der Fall sein. Durch eine präzise Registrierung können die Einzelbilder einer Bildfolge zu einem Bildteppich verrechnet und somit ein großes, statisches Einzelbild generiert werden (vgl. Abbildung 45), (Müller, et al., 2001). Das Verfahren arbeitet multisensoriell, wie an den Registrierungsergebnissen einer Zweifarb-IR-Kamera (MWIR und LWIR) zu sehen ist (Müller, et al., 2007), vgl. Abbildung 46.

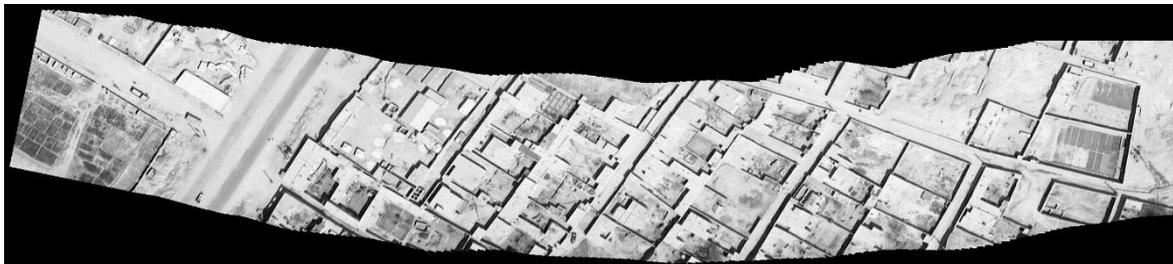


Abbildung 45: Zu beachten ist, dass bei dem Bildteppich keine Schnittkanten erkennbar sind, was an der hohen Subpixelgenauigkeit ($< 0,2$ Pixel) liegt. Bild: IOSB.

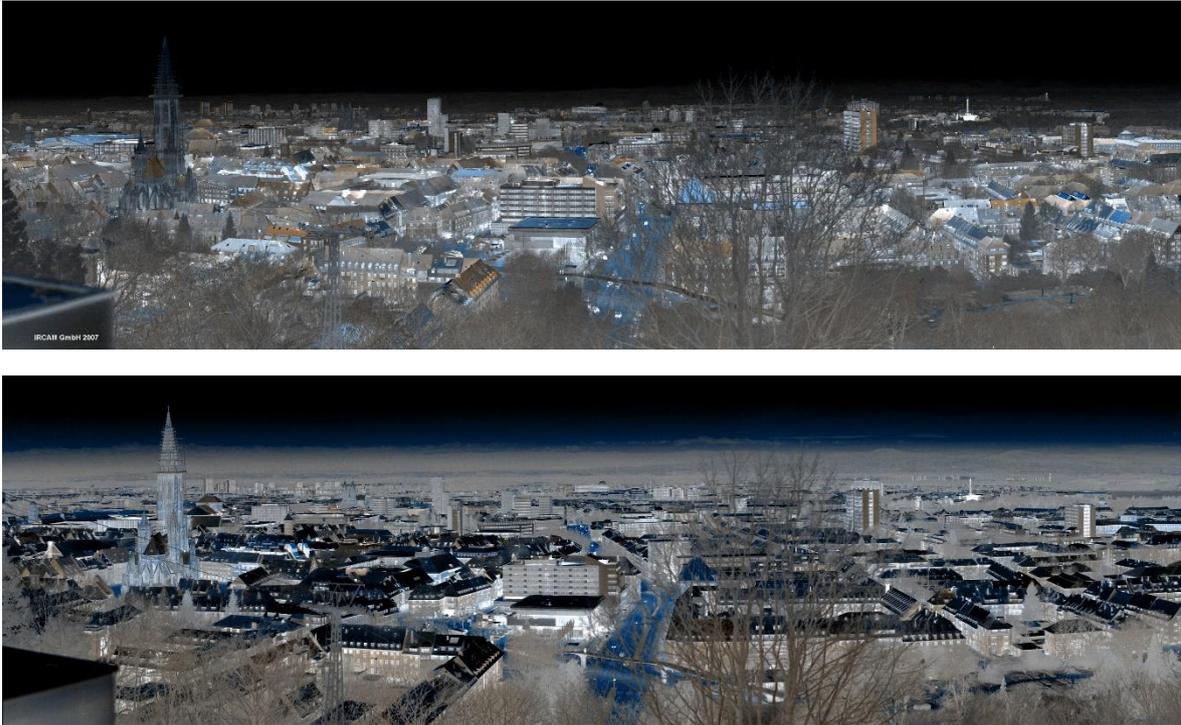


Abbildung 46: Bildteppiche einer Zweifarb-IR-Kamera (Auflösung: 3.000 x 900 Pixel) von Freiburg. Oben: Tagaufnahme, unten: Nachtaufnahme. Die MWIR- und LWIR-Bänder sind mit Komplementärfarben überlagert. Verwendet wurden rund 100 Einzelbilder. Einzelbilder: IRCAM GmbH; Bildteppich: IOSB.

Der generierte Bildteppich kann auf vielfältige Weise weiterverwendet werden.

- Referenzierung mit einer geografischen (Vektor-) Karte oder einem Referenzbild (z.B. einem Orthofoto (vgl. Abbildung 47) oder einem anderen Bildteppich). Damit sind u.a. die grundlegenden Voraussetzungen zur Änderungsdetektion geschaffen. Bei der Referenzierung zweier Bildteppiche aufeinander oder eines Bildteppichs auf eine Bildfolge fungiert der eine Bildteppich als szenarischer Soll-Zustand und der andere referenzierte Bildteppich bzw. die referenzierte Bildfolge als Ist-Zustand.
- Die Referenzierung mit einer Karte erlaubt in Folgeschritten eine nachträgliche Verbesserung der Geocodierungsqualität, also der weltlichen Koordinatenangabe. Systeme ohne diese Funktion verlassen sich ausschließlich auf die ungenauen Flug- und Aufnahmeparameter. Damit sind jedoch keine präzisen Ortsangaben in Weltkoordinaten möglich, um z.B. Einsatzkräfte aus der Luft gezielt dirigieren zu können.



Abbildung 47: Bildteppich, referenziert mit einem Ikonos-Orthobild (Hintergrund) und in blau eingezeichnetem Flugweg des Sensorträgers (gemäß GPS-Flugdaten). Bild: IOSB, Ikonos-Bild: SGA²¹/Gelsdorf.

- Aufgrund der hohen Überlappung der Einzelbilder eines Bildteppichs bietet sich eine Stereobildteppichgenerierung (Abbildung 48) an, aus der (derzeit noch nicht in Echtzeit) ein Tiefenbildteppich generiert werden kann. Abbildung 49 zeigt einen Ausschnitt aus einem Beispiel für einen Tiefenbildteppich.



Abbildung 48: Beispiel für einen Stereobildteppich. Bild: IOSB.

²¹ SGA: Satellitengestützte Aufklärung

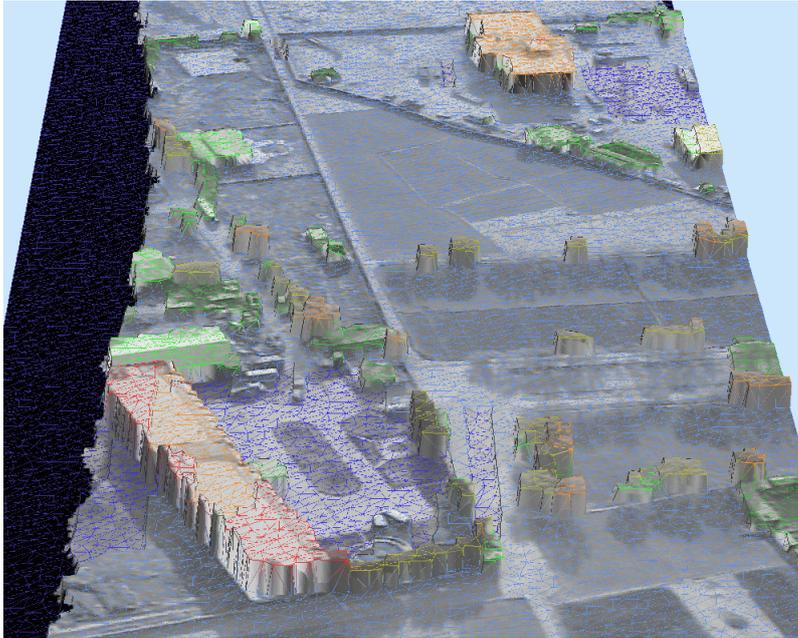


Abbildung 49: Beispiel für einen Tiefenbildteppich. Bild: IOSB.

Die Einbindung der Bildteppichfunktionalität (wie auch diverser anderer Funktionen) in das System ABUL erfolgt in einem zweiten Fenster (Abbildung 50).

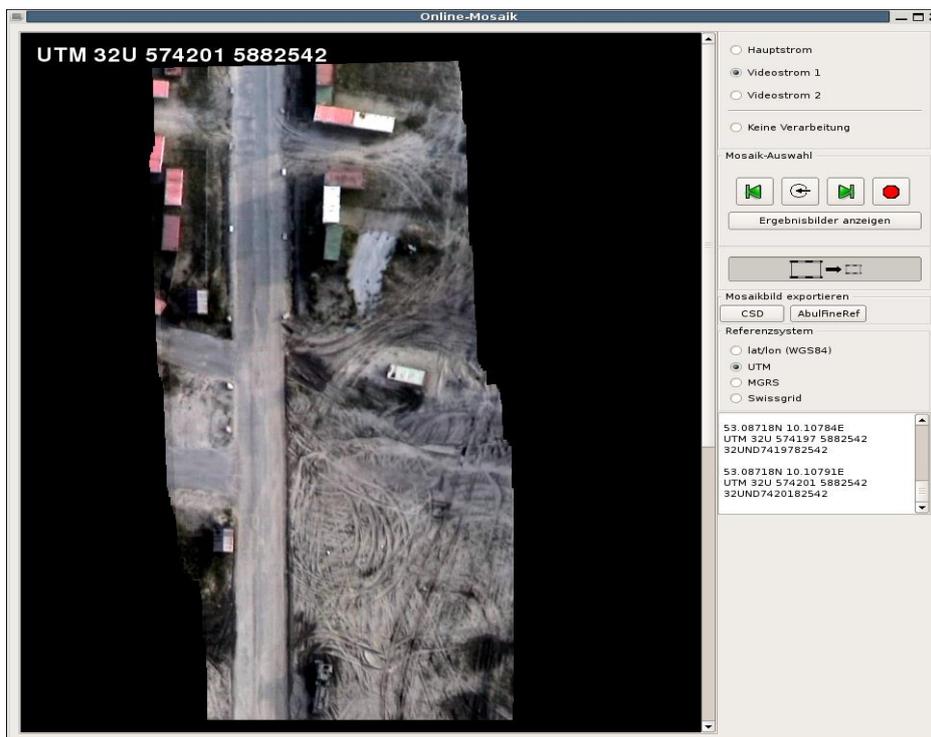


Abbildung 50: In Echtzeit generiertes Bildteppichergebnis. Bild: IOSB.

7.3.3 Bildverbesserung und Bildstabilisierung

Eine präzise Bildregistrierung erlaubt fortschrittliche Bildkorrekturen, wie z.B. die Entfernung von Sensorartefakten und Pixelstörungen (engl. *dead pixels*), vgl. (Heinze, et al., 2008).

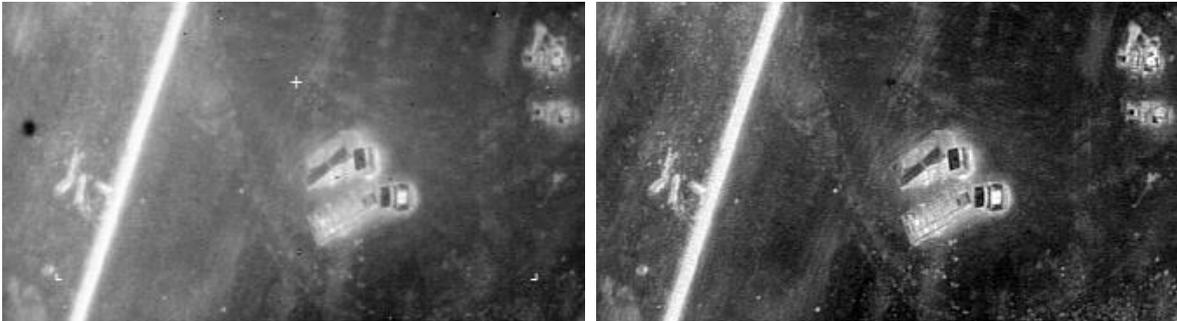


Abbildung 51: Bildverbesserung (u.a. Beseitigung von Pixelstörungen und Kontrastanhebung). Original: links, Korrektur: rechts. Bild: IOSB.

Aus preislichen oder systemtechnischen Gründen (Gewicht, Größe) ist oft keine mechanische Stabilisierung des Sensors möglich bzw. erwünscht. Andererseits erlauben die Flug- und Aufnahmeparameter aufgrund ihrer Ungenauigkeit keine darauf basierende Verrechnung. Daher sind auf der Basis der Referenzierungsalgorithmik Softwarelösungen realisiert worden, die eine stufenlos parametrierbare Stabilisierung ermöglichen (Heinze, et al., 2008).

7.3.4 Weitere Verfahrensansätze

Die Interaktion mit Bildfolgen in Echtzeit fordert vom Benutzer viel Fingerspitzengefühl. U.a., um Zielobjekte zuverlässiger und schneller markieren zu können, laufen Arbeiten, die sich auf die Erfassung der Augenbewegungen des Benutzers beziehen (Hild, et al., 2013). Damit soll der Benutzer in der Lage sein, das ihn interessierende, ggf. sich bewegende Objekt in der Bildfolge mit seinen Augen anzuvisieren und mit einer zusätzlichen Interaktion (Tastendruck) zu markieren.

Aufgrund der Möglichkeit zur Schätzung der Eigenbewegung durch die Bildreferenzierung ist eine Grundlage geschaffen, die Eigenbewegung aus dem Bildstrom heraus zu rechnen und sich relativ zum Hintergrund bewegende Objekte zu detektieren

(bildgestützte Bewegzieldetektion – *image-based moving target indication*). Ergebnisse zeigt die folgende Abbildung, entsprechende Arbeiten wurden u.a. in (Teutsch, et al., 2011) beschrieben.



Abbildung 52: Bildgestützte Bewegzieldetektion. Bild: IOSB.

Zur Verwaltung und leichteren Abrufbarkeit erfolgter Sensordaten ist ein (Video-) Archivsystem nötig. Unter der Bezeichnung »VABUL« sind entsprechende Arbeiten im Gange. Die Arbeiten zielen u.a. auf ein bildinhaltsgestütztes, also semantisches Retrieval von Bilddaten ab (*image retrieval*). Dabei soll z.B. ein Ausschnitt aus einer Szene vorgegeben werden, um die Videos in der Datenbank nach den Stellen zu durchmustern, die Aufnahmen von der gesuchten Szene darstellen. Die gegenwärtig realisierte Archivierungsfunktionalität erlaubt es u.a., interaktiv eine geographische Region zu spezifizieren, um zeitlich geordnet die diesbezüglichen erfolgten Daten abzurufen (Brüstle, et al., 2012).

8 Zusammenfassung und Ausblick

In dieser Arbeit wurden Beiträge für Sicherheitsanwendungen aufgezeigt und exemplarisch angewendet. Im Wesentlichen wurde ein flexibler Ansatz zur Auffälligkeits-/Saliendetektion in Kombination mit Aspekten einer schnellen SIMD-Implementierung dargestellt und ein Bildfolgenauswertesystem ABUL präsentiert.

Der Ansatz zur Auffälligkeitsdetektion ist flexibel und leicht erweiterbar, da die gesuchten Strukturen in Form einfacher Binärmodelle vorgebar sind. Es reicht also, die grundlegenden Eigenschaften der jeweiligen Auffälligkeit in Form einer binärbildlichen Skizzierung zu modellieren. Jedes Modell kann einen Softwareagenten realisieren, der dann im Bildbereich nach den entsprechenden Mustern sucht und anderen Auswerteverfahren Hinweise auf Orte liefert, die auf der Basis anderer Binärmodelle (oder grundsätzlich anderer Verfahren) inspiziert werden sollten.

Ausprägungen des ABUL-Systemansatzes sind in der Praxis von diversen behördlichen Stellen in Erprobung bzw. im Einsatz. Die Entscheidung, einen solchen Systemansatz »von Null auf« zu entwickeln, barg das Risiko der Finanzierbarkeit, welche jedoch durch die Akquisition entsprechender Mittel in einem Maße erfüllt werden konnte, dass mittlerweile zwei Forschungsgruppen – in Teilen – an der Weiterentwicklung beteiligt sind.

Die ersten Funktionen bezogen sich auf den primär interaktiven Umgang mit Bildfolgen, wobei bereits in den frühen Versionen echtzeitfähige Bildregistrierungsverfahren integriert wurden. Diese Verfahren wurden als grundlegend für eine Vielzahl weiterer Anwendungsmöglichkeiten, insbes. zur Auffälligkeits-/Änderungsdetektion, erkannt und ihre kontinuierliche Fortentwicklung stetig weiterbetrieben.

Unmittelbar naheliegende Funktionserweiterungen bieten sich z.B. wie folgt an:

- Registrierung einer Einzelbildfolge auf einen zuvor generierten (Referenz-) Bildteppich. Die Referenz stellt einen szenarischen Soll-Zustand dar, der mit dem Ist-Zustand der live und ggf. »aus der Hand« aufgenommenen Bildfolge im Hinblick auf Auffälligkeiten/Veränderungen verglichen werden soll.

- Falls k Sensoren an das System angeschlossen sind, bietet sich die Erstellung von k -Bildteppichen an. Diese können jeweils einzeln für sich erzeugt werden, um dann eine gegenseitige Referenzierung zu erfahren.
- Auf der Basis von Stereo-Bildteppichen sind Verfahren wünschenswert, die entsprechende Tiefeninformationen extrahieren und daraus einen »Tiefenbildteppich« generieren. Erste Ansätze sind vorhanden, jedoch sind sie noch recht rechenintensiv, sodass noch kein Echtzeitansatz möglich ist.
- Sind Tiefenbildteppiche gegeben, bietet sich eine Texturierung entsprechend einer Auswahl aus den k -Bildteppichen an.

9 Anhang NAG-Herleitung

In diesem Anhang soll die »Normierte Ableitung einer Gaußfunktion« (NAG) auf alternative Weise hergeleitet und einige Eigenschaften dargelegt werden. Diese Herleitung der NAG-Funktion demonstriert insbesondere eine alternative Behandlung des Skalierungsparameters.

Die Nutzung der Ableitung einer Gaußfunktion zur Kantendetektion bzw. Detektion von Konturpunktkandidaten wurde parallel und unabhängig voneinander von A. KORN und J. F. CANNY in ihren Veröffentlichungen (Korn, 1985), (Korn, 1988) bzw. (Canny, 1983), (Canny, 1986) vorgeschlagen. Jeweils ergänzende Aspekte beziehen sich auf Kriterien guter Kanten (Canny, 1986) bzw. der Normierung der Gauß-Ableitung für die Betrachtung im Skalenraum (Korn, 1988).

Die hier dargestellte Herleitung ist eine Variante der in (Korn, 1988) vorgenommenen Herleitung. Im Gegensatz zum Originalpapier wird jedoch auf die Einführung einer eigenen Konstante für die Normierung verzichtet und die Darstellung gestrafft. Auch ist der Denkansatz für die Herleitung etwas anders bzw. allgemeiner gehalten: es geht letztlich um die Bestimmung von Formparametern einer Gaußfunktion bzw. deren Ableitung. Und zwar in beiden Fällen mit dem Ziel, Dichtefunktionen zu erhalten. Es wird sich zeigen, dass

- a. die Eigenschaft, Dichtefunktion zu sein, zwangsläufig notwendig ist, um eine Gaußfunktion (bzw. deren Ableitung oder einer beliebigen anderen Funktion) für Betrachtungen für Filterbänke (Skalenräume oder Auflösungspyramiden) verwenden zu können und
- b. auf die Einführung zusätzlicher Formparameter zur Normierung der Ableitung einer Gaußfunktion (zu einer links- bzw. rechtsseitigen Dichtefunktion) verzichtet werden kann.

Es wird aufgrund der Separierbarkeit der Gaußfunktionen zur Herleitung nur der eindimensionale Fall betrachtet. Das Ergebnis für den zweidimensionalen Fall wird in Abschnitt 9.1 gezeigt.

C. F. GAUß schlug zur Messfehlerbeschreibung folgende Funktionsterme vor:

$$g(x) = c \cdot e^{-ax^2}, \quad (27)$$

wobei $c, a \in \mathbb{R}^{>0}$ Formparameter sind. Zur Ermittlung der Differenz mittlerer Grauwerte benachbarter Gebiete schlug (Korn, 1988) die Faltung der Bildfunktion f mit g' vor:

$$g' = -2a \cdot x \cdot c \cdot e^{-ax^2}, \quad (28)$$

was sich zu

$$g' = -2a \cdot x \cdot g(x) \quad (29)$$

vereinfachen lässt.

Um Ergebnisbilder verschiedener Auflösungsstufen vergleichen zu können, wurde von KORN ferner ein Faktor k vorgeschlagen, der die Summe der Filterwerte auf ± 1 normieren soll, d.h. für den rechtsseitigen Fall (linksseitig analog):

$$k \int_0^{\infty} g'(x) dx = 1 = k \lim_{x \rightarrow \infty} g(x) - g(0) = k \cdot -c \rightarrow k = \frac{1}{c} \quad (30)$$

Für die Standardabweichung σ erhält man bei einer Normierung von g zu einer Dichtefunktion

$$\int_{-\infty}^{\infty} g(x) dx = 1 \quad (31)$$

die Formparameter $a = \frac{1}{2\sigma^2}$ und $c = \frac{1}{\sigma\sqrt{2\pi}}$. Da g' punktsymmetrisch zum Ursprung ist, wird verlangt, dass die jeweiligen Beträge der links- bzw. rechtsseitigen Teile der Funktion Dichtefunktionen darstellen. Für den rechtsseitigen Teil (linksseitiger Fall analog) also, dass:

$$k \int_0^{\infty} g'(x) dx = 1. \quad (32)$$

zu erfüllen ist. Der Normierungsfaktor k kann aus den Rechenregeln für unbestimmte Integrale berechnet werden (33).

$$\int_0^{\infty} g'(x) dx = \lim_{x \rightarrow \infty} g(x) - g(0) = -g(0) = -c \quad (33)$$

Somit ergibt sich auch hier: $k = \frac{1}{c} = \sigma\sqrt{2\pi}$.

Für die gewünschte Normierung der ersten Ableitung einer Gaußfunktion genügt es also, die Ableitung g' mit dem Kehrwert des für g bereits festgelegten Formparameters c zu multiplizieren:

$$NAG(x) = \frac{1}{c} g'(x) = \frac{1}{c} -2a \cdot x \cdot g(x) = \frac{1}{c} -2a \cdot x \cdot c \cdot e^{-ax^2} = -2a \cdot x \cdot e^{-ax^2}, \quad (34)$$

erneut mit $a = \frac{1}{2\sigma^2}$.

Wie zu erkennen ist, fällt der Formparameter c bei der NAG-Berechnung ganz heraus und die Berechnung vereinfacht sich. Erkennbar ist ferner, dass die gewünschte NAG auf kanonische Weise unter Nutzung bereits für g bestimmter Formparameter herleitbar ist, d.h. weitere Form- bzw. Normierungsparameter sind nicht erforderlich.

Die Formel für die NAG skaliert korrekt im kontinuierlichen Bereich. Zur praktischen Berechnung von diskreten Maskenkoeffizienten ist sie jedoch eher ungeeignet, da zwangsläufig Fehler (d.h. Summe der positiven bzw. negativen Koeffizienten $\neq 1$ bzw. $\neq -1$) aufgrund der endlichen Anzahl von Maskenkoeffizienten auftreten. Die diskrete NAG muss also »nachnormiert« werden. Anstatt die NAG bei der tatsächlichen Koeffizientenberechnung einer Nachnormierung zu unterziehen, können die Koeffizienten gemäß zweiter Ableitung von g berechnet werden. Die Normierung ergibt sich dann durch die Multiplikation mit dem Mittelwert der positiven Koeffizienten.

9.1 Zweidimensionale Berechnung der NAG

Die Amplituden und Richtungen der Gradienten werden mit (35) berechnet.

$$NAG_{\sigma}(x, y) = \frac{1}{c_{\sigma}} g_{\sigma}(x, y) = -2a_{\sigma} \cdot x \cdot e^{-a_{\sigma} x^2}. \quad (35)$$

Der Skalierungsparameter σ wurde als Index in (35) ergänzt. Die Korn'sche Transformation K der Bildfunktion f gemäß einem eingestellten σ ergibt sich für die x - und y -Richtung aus (36).

$$K_{\sigma}(x, y) = f(x, y) * \begin{pmatrix} NA G_{\sigma}^x(x, y) \\ NA G_{\sigma}^y(x, y) \end{pmatrix}. \quad (36)$$

Nach den Faltungen liegen zwei Ergebnisse vor: K_{σ}^x und K_{σ}^y . Die Amplitude A der Gradienten (Kontrastmaß) errechnet sich aus (37).

$$A(x, y) = \sqrt{K_{\sigma}^x(x, y)^2 + K_{\sigma}^y(x, y)^2} \quad (37)$$

und der Gradientenwinkel φ bestimmt sich durch (38):

$$\varphi(x, y) = \arctan \left(\frac{K_{\sigma}^x(x, y)}{K_{\sigma}^y(x, y)} \right). \quad (38)$$

Abbildung 53 zeigt eine grafische Darstellung einer zweidimensionalen NAG.

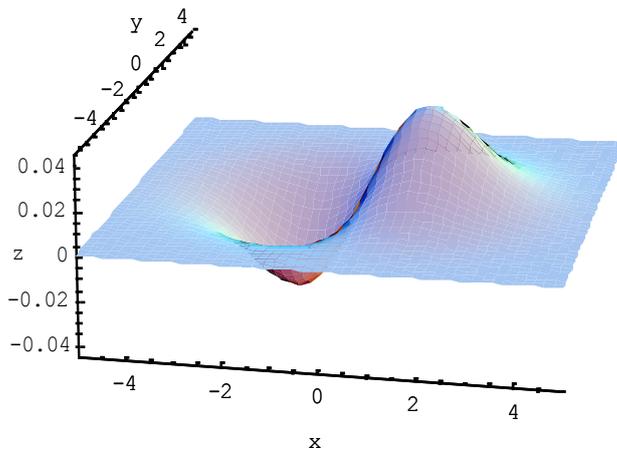


Abbildung 53: Die zweidimensionale NAG-Funktion ($\sigma = 2$).

10 Eigene Erfindungen / Patente

Im Folgenden sollen zwei Geräte vorgestellt werden, deren Prüfung auf Patentierungsfähigkeit mit dem Vorlegen dieser Arbeit angestoßen wird. Beide Geräte dienen der Umfelderkennung für inspizierende Aufgaben.

Beide Geräte stellen Sensorverbauungen zur Umfelderkennung, z.B. für Sicherheitsaufgaben, dar. Dabei liegt der Schwerpunkt Fällen auf einer möglichst großflächigen Erfassung des Umfelds, ohne Kompromisse bei der Auflösung eingehen zu müssen.

10.1 360°-Panoramakamera

Die heutigen Techniken zur Erstellung eines 360°-Horizontalpanoramas lauten wie folgt:

- Vorschaltung eines 360°-Spiegels (Kegelspiegel) vor das Objektiv. Der Nachteil besteht in dem deutlichen Auflösungsverlust, denn die zur Verfügung stehende Sensorfläche verteilt sich auf die kompletten 360°.
- Rotierende Kamera auf einem entsprechenden Stativkopf. Der Nachteil besteht in einer aufwändigen und damit teuren Elektromechanik.
- Zeilenkameras, die durch die Bewegung der Kamera oder der Objektivrichtung zeilenweise das 360°-Panorama generieren. Nachteil: teurer Aufbau.
- Wiederholte Einzelaufnahmen mit einer gängigen Flächenkamera (aus der Hand oder mit einem Stativ). Die Einzelaufnahmen werden dabei mit Hilfe entsprechender Software nachbearbeitet und zum gewünschten Panoramabild zusammengesetzt. Nachteil: Ergebnis liegt erst nach der Nachbearbeitung vor, der gesamte Prozess ist aufwändig.
- Sich ggf. drehende Laser-Scanner, die ihre Umgebung vermessen. Z.B. Faro-Laserscanner. Allerdings benötigt so ein Scanner rund 8 Minuten für einen Scan und darf nicht bewegt werden.

Zur Kompensation der Einzelnachteile wird ein System zur Panoramaerzeugung vorgeschlagen, das im Wesentlichen aus zwei Komponenten besteht: a) einem rotierenden, schräg gestellten Spiegel über einer feststehenden Kamera, beides in ein entsprechendes Gehäuse eingebaut sowie b) der in Abschnitt 7.3. vorgestellten Bildregistrierungsverfahren. Auch die Bildregistrierungsverfahren sind entscheidend, denn das mit dem rotierenden Spiegel erfasste Bild vollzieht seinerseits bei einem Spiegelumlauf eine 360°-Drehung (Abbildung 54).



Abbildung 54: Einzelbilder aus einer Rundumerfassung der 360°-Panoramakamera.

Die gedrehten Einzelbilder der Bildfolge müssen also in die aufrechte Position gedreht und dann aufeinander registriert werden, um das 360°-Panorama zu erzeugen.



Abbildung 55: Beispielhafte Veranschaulichung der Drehung von Einzelbildern.

In Abbildung 56 ist ein Prototyp gezeigt. Er besteht aus einem unteren Aluminiumkorpus, der die vertikal nach oben ausgerichtete Kamera beherbergt. Ferner einen darüber angebrachten Glaskörper, in dem sich der über dem Kameraobjektiv rotierende Spiegel befindet. In dem Glaskörper sind auch Induktionssensoren sichtbar, die die Erkennung des Start-/Endpunktes einer Umdrehung ermöglichen. Schließlich

findet sich noch ein oberer Aluminiumkorpus, der den Motor zum Drehen des Spiegels enthält.

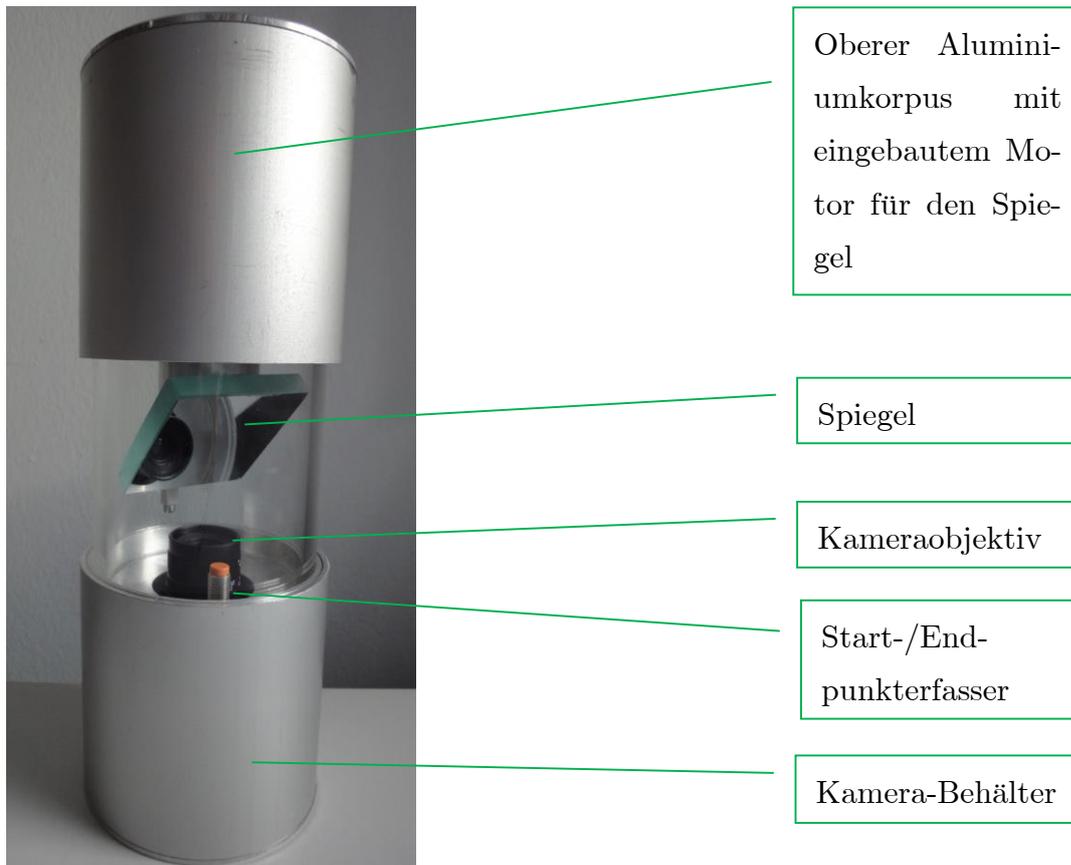


Abbildung 56: Prototyp der Rundumkamera.

10.2 Scanner-Tracker

Eine weitere Erfindung bezieht sich auf die Anpassung eines Geräts, welches aus der Lichttechnik bekannt ist. Auch diese Adaption soll, wie die Rundumkamera, einen möglichst großen Bereich erfassen können, ohne einen Auflösungsverlust hinzunehmen. Aufgrund der von Haus aus eingebauten und standardisierten Digitalschnittstelle (MIDI) eignet sich das Gerät insbesondere für die Verfolgung von sich bewegenden Objekten mit großflächigem Erfassungsbereich.

10.2.1 MIDI

MIDI steht für »Musical Instrument Digital Interface« und stellt eine industriell standardisierte (Hardware & Protokoll), digitale Schnittstelle zum Austausch von Steuerinformationen zwischen elektronischen Geräten dar. Ursprünglich wurde es für die digitale Steuerung von entsprechend ausgerüsteten Musikinstrumenten (z.B. Synthesizern) entwickelt, wurde jedoch auch auf Effekt- und Aktuatorikgeräte erweitert. Zusätzlich zu den Steuerungsmöglichkeiten sind auch Möglichkeiten zur zeitlichen Synchronisation vorgesehen:

a) MIDI-Clock

Basiert auf Takt- und Notenebene.

b) MIDI-Time-Code

Dieser Synchronisationsmodus basiert auf dem SMPTE-Timecode (Society of Motion Picture and Television Engineers, New York), womit im professionellen Video-, TV- und Studiobereich beliebige Video- und Audiotechniken miteinander synchronisiert werden können. Der Code basiert auf einer absoluten Zeitangabe. Der SMPTE-Verband bestimmt maßgeblich u.a. auch den HDTV²²- und SDI²³-Standard seit seinen Anfängen bis heute.

10.2.2 Moving-Light-Geräte

Eine Rundumleuchte strahlt Licht über einen Bereich von 360° ab. Dabei handelt es sich meist um Drehspiegelleuchten, die via Spiegel einen Lichtkegel in die Szene projizieren. Als Warn- und Erkennungszeichen sind sie insbesondere im BOS²⁴-Bereich in Verwendung.

²² HDTV: High Definition Television

²³ SDI: Serial Digital Interface – Übertragung unkomprimierter Videodaten über Koax- oder Lichtwellenleiter

²⁴ BOS: Behörden und Organisationen mit Sicherheitsaufgaben

Es gibt zwei grundlegende Arten von Moving-Lights im Bereich der Lichttechnik: zum einen sog. »Scanner«, bei denen der austretende Lichtstrahl durch einen elektromechanisch zweiachsig beweglichen Spiegel abgelenkt wird (Abbildung 57) und sog. »Moving-Heads«, bei denen die komplette Projektionseinheit bewegt wird. Ein Scanner besitzt den Vorteil der schnelleren und genaueren Positionierbarkeit aufgrund des leichtgewichtigen Ablenkspiegels, wohingegen ein Moving-Head über einen größeren Erfassungsbereich verfügt (je nach Umsetzung mehr als einen Halbkreis, bei einem Scanner lediglich rund $160^\circ \times 30^\circ$). Bei beiden Systemen sind alle Funktionen über eine MIDI-Schnittstelle fernsteuerbar.

Es liegt die Idee nahe, anstelle des Leuchtmittels eine Kamera einzubauen. Damit geht es also nicht mehr darum, einen Lichtstrahl in die Szene zu werfen, sondern das Licht von bestimmten Szenenbereichen einzufangen. Folgende Vorteile weist ein solches elektromechanisch positionierbares Sensorsystem auf:

1. Nach jeder Positionierung wird das Licht mit voller Kameraauflösung eingefangen.
2. Kommerziell erhältliche Scanner sind für professionelle Ansprüche konzipiert und erlauben auch den Dauereinsatz.
3. Durch die bereits integrierte MIDI-Schnittstelle ist eine Ansteuerung durch entsprechende Rechnerschnittstellen bzw. Steuersoftware gegeben.
4. Die Geräte sind kompakt und leicht transportierbar.

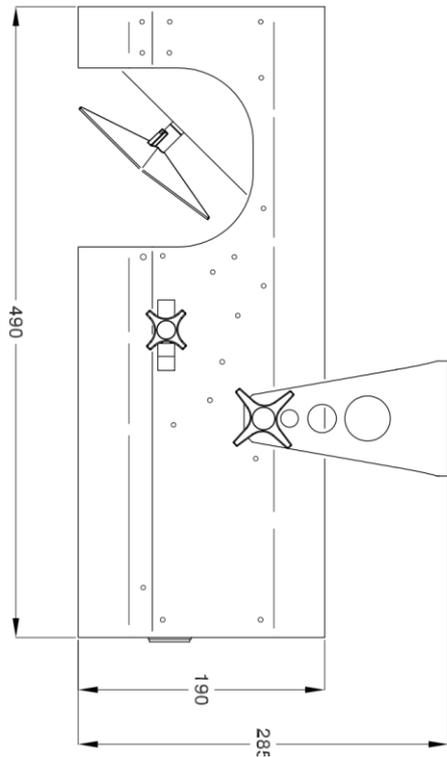


Abbildung 57: Seitenansicht eines typischen Scanners (Bild: Firma Martin).

Zusammen mit einer haus-eigenen Software zur Detektion und Verfolgung von Personen in Bildsequenzen, z.B. (Monari, 2011) oder, auf IR bezogen (Jüngling, et al., 2011), lässt sich mit dem Scanner-Tracker (Abbildung 59) eine entsprechende Anwendung für den Sicherheitsbereich realisieren (Abbildung 58).

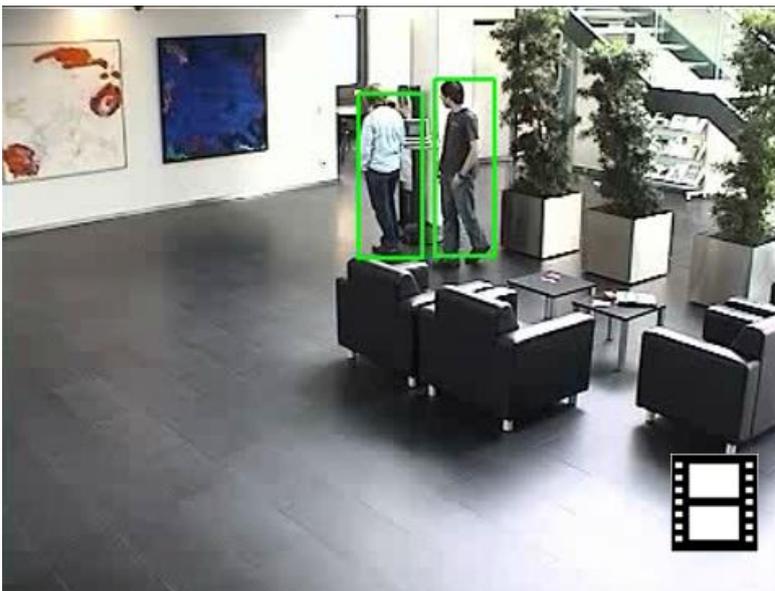


Abbildung 58: Personendetektion und Tracking (Bild: IOSB).

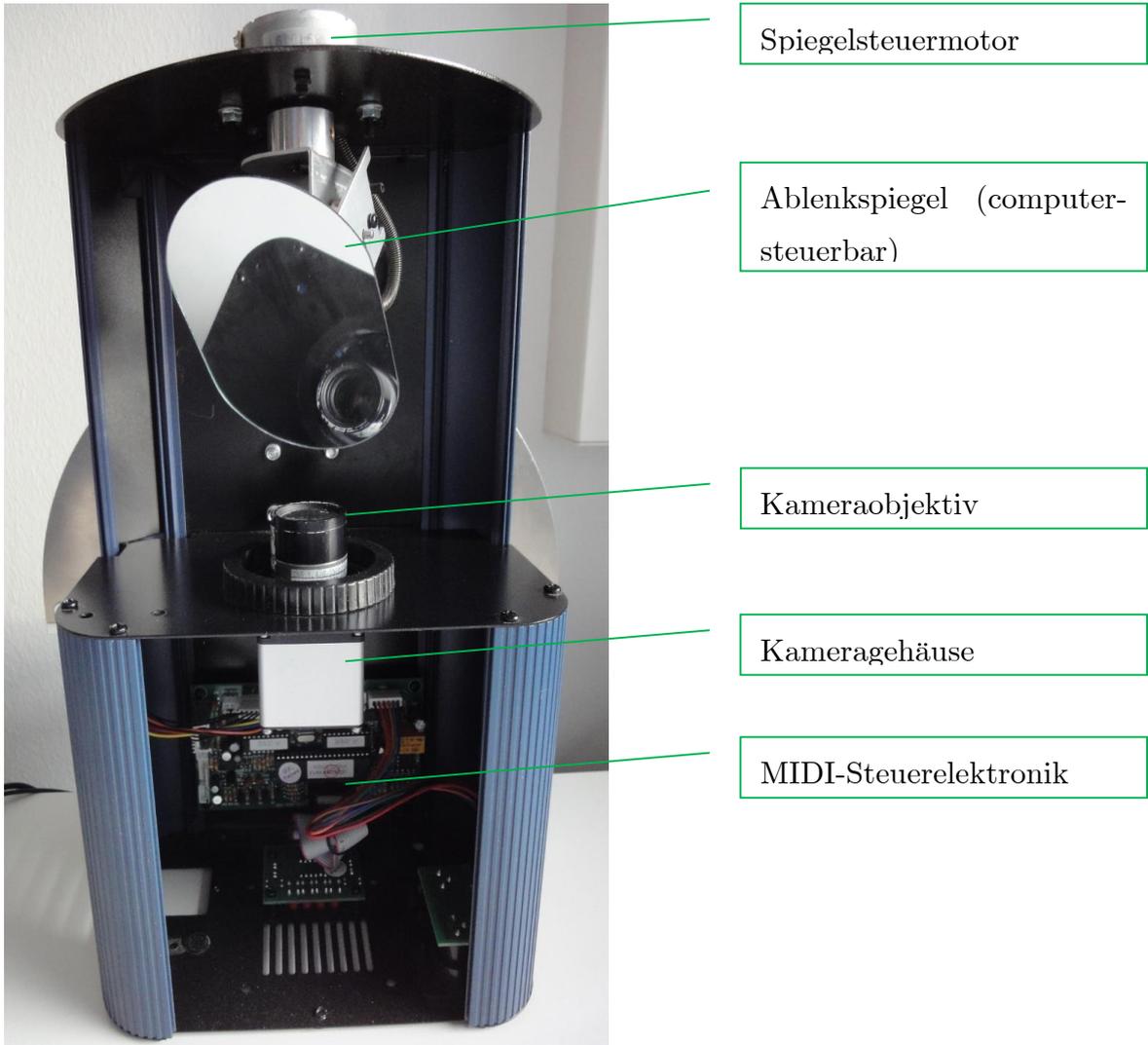


Abbildung 59: Prototyp des Scanner-Tracker.

11 Eigene Publikationen

T. Müller und M. Müller:

CART: Computer Aided Camouflage Assessment.

Internationales Symposium für Indirekte Schutzsysteme ISIS 2012, October 8. - 12., 2012, Bad Reichenhall, Germany.

Müller, Thomas; Müller, Markus:

CART V: Recent advancements in computer aided camouflage assessment.

In: Holst, Gerald. C. (Ed.); Krapels, Keith A. (Ed.): SPIE Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXI: 26.04.2011, Orlando, Florida, USA. Bellingham, Washington: SPIE - The International Society for Optical Engineering, 2011, 11 S. (Proceedings of SPIE 8014).

Schreiber, Ralf; Breiter, Rainer; Ziegler, Johann; Cabanski, Wolfgang; Müller, Markus; Heinze, Norbert; Walther, Martin; Rehm, Robert:

Multi-spectral and hyperspectral IR-sensors for improved surveillance applications.

In: Fraunhofer Alliance for Defense and Security Research VVS; Fraunhofer-Verband Verteidigungs- und Sicherheitsforschung VVS: Future Security: 6th Security Research Conference, Berlin, September 7th - 9th, 2011. Stuttgart: Fraunhofer Verlag, 2011, S. 321-325.

Müller, Thomas; Müller, Markus:

Vision-based drone flight control and crowd or riot analysis with efficient color histogram based tracking.

In: Henry, Daniel J. (Ed.); Society of Photo-Optical Instrumentation Engineers (SPIE), Bellingham, Wash.: SPIE Defense, Security and Sensing 2011: Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications VIII: 27 April 2011, Orlando, Florida, United States. Bellingham, Wash.: SPIE, 2011, 14 S. (Proceedings of SPIE 8020).

Müller, Thomas; Müller, Markus:

CART IV: improving automatic camouflage assessment with assistance methods.

In: Holst, Gerald C. (Ed.); Society of Photo-Optical Instrumentation Engineers

(SPIE), Bellingham, Wash.: SPIE Defense, Security and Sensing 2010: Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXI: 6 April 2010, Orlando, Florida, United States. Bellingham, Wash.: SPIE, 2010, 11 S. (Proceedings of SPIE 7662).

T. Müller und M. Müller:

Camouflage Assessment for Moving Targets (CART III).

Internationales Symposium für Indirekte Schutzsysteme ISIS 2009, Bad Reichenhall, 20. - 23.10.2009.

Müller, Thomas; Honke, Thomas; Müller, Markus:

CART III: improved camouflage assessment using moving target indication.

In: Holst, Gerald C. (Ed.): Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XX: Orlando, FL, 14-16 April 2009. 2009, 12 S. (Proceedings of SPIE 7300).

Müller, Markus:

NEST: Aktiver Feldlager- und Liegenschaftsschutz.

In: Strategie und Technik: Streitkräfte, Rüstung, Sicherheit. 52 (2009), 9, S. 68-69.

Müller, Thomas; Müller, Markus:

CACAMO - computer-aided camouflage assessment of moving objects.

In: Holst, Gerald C. (Ed.): SPIE Defense and Security 2008: Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XIX. 2008 (Proceedings of SPIE 6941).

Emter, T.; Monari, E.; Frey, Ch. W.; Müller, T.; Kuntze, H.-B.; Laubenheimer, A.; Müller, M.:

AMROS - ein autonomes mobiles Robotersystem zur multisensoriellen Überwachung von Liegenschaften.

In: Beyerer, Jürgen (Ed.); Fraunhofer Verbund für Verteidigungs- und Sicherheitsforschung: Future security 2007: 2nd Security Research Conference: 2007, 12th - 14th September. Karlsruhe : Universitätsverl. Karlsruhe, 2007, S. 151-154.

Müller, Thomas; Müller, Markus:

Computer-aided camouflage assessment in real-time.

In: Holst, Gerald C. (Ed.): Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVIII: Orlando, FL, 11 April 2007. 2007 (Proceedings of SPIE 6543).

Müller, Markus; Schreer, Oliver; Lopez Saenz, Monica:

Real-time image processing and fusion for a new high-speed dual-band infrared camera.

In: Holst, Gerald C. (Ed.): Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVIII: Orlando, FL, 11 April 2007. 2007, 9 S. (Proceedings of SPIE 6543).

Müller, Markus; Krüger, Wolfgang; Saur, Günter:

Robust image registration for fusion.

In: Information fusion. 8 (2007), 4, S. 347-353.

Müller, Markus:

Bewegte Video- und IR-Sensorik: Mehrwert durch aktuelle Bildauswertungsverfahren.

In: Homeland Security. (2005), 2, S. 60-61.

Krüger, W.; Müller, M.:

m³motion - Detektion und Verfolgung von Objekten bei bewegter Kameraplattform.

In: Wirth, Sibylle (Red.); Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB), Karlsruhe: Fraunhofer-Institut Informations- und Datenverarbeitung IITB: Jahresbericht 2004. Karlsruhe, 2005, S. 22-23.

Müller, Markus; Saur, Günter ; Krüger, Wolfgang:

Real-time exploitation of image sequences in dynamic surroundings.

In: Watkins, Wendell R. (Eds.); Society of Photo-Optical Instrumentation Engineers (SPIE), Bellingham, Wash.: Targets and backgrounds XI: characterization and representation. 28 - 29 March 2005, Orlando, Fl. Bellingham, Wash.: SPIE, 2005, p. 133-137. (Proceedings of SPIE 5811).

Müller, M.; Heinze, N.:

Automatische Bildauswerteverfahren im interaktiven Systemverbund.

In: Wirth, Sibylle (Red.) ; Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB), Karlsruhe: Fraunhofer-Institut Informations- und Datenverarbeitung IITB: Jahresbericht 2003. Karlsruhe, 2004, S. 36-37.

Müller, Markus; Heinze, Norbert; Clement, Dieter:

ATR-based camouflage effectiveness evaluation of MUSTAFA targets.

In: Watkins, Wendell R. (Eds.); Society of Photo-Optical Instrumentation Engineers (SPIE), Bellingham, Wash.: Targets and backgrounds IX: Characterization and visualization. Orlando, Florida, 21-22 April 2003. Bellingham, Wash.: SPIE, 2003, S. 162-168. (Proceedings of SPIE 5075).

Müller, Markus; Heinze, Norbert:

Das dritte Auge: Interaktion mit automatischen Bildauswerteverfahren.

In: visIT, Fraunhofer IITB. (2003), 2/2003, S. 8-9.

Müller, Markus; Zimmermann, Georg:

Automatische Stereobildgenerierung.

In: Wirth, Sibylle (Red.) ; Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB), Karlsruhe: Fraunhofer-Institut Informations- und Datenverarbeitung IITB: Jahresbericht 2001. Karlsruhe, 2002, S. 34-35.

Müller, Markus; Krüger, Wolfgang; Heinze, Norbert:

Multisensor data fusion and GIS utilization for ATR.

In: Watkins, Wendell R. (Eds.); Society of Photo-Optical Instrumentation Engineers (SPIE), Bellingham, Wash.: Targets and backgrounds VII: Characterization, visualization and the detection process. Orlando, Florida, 16-17 April 2001. Bellingham, Wash.: SPIE, 2001, S. 36-41. (Proceedings of SPIE 4370).

Link, Norbert; Müller, Markus; Krüger, Wolfgang; Heinze, Norbert:

ATR components for aerial image analysis.

In: University of Bristol: UAVs: Unmanned air vehicle systems. 15th international conference, Bristol, UK, 10-12 April 2000: Conference papers. Bristol, 2000, S. 27.1-27.9.

Müller, Markus ; Heinze, Norbert:

Automatische Klassifikation von Hautzuständen.

In: Brödner, Ute (Red.); Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB), Karlsruhe: Fraunhofer-Institut Informations- und Datenverarbeitung IITB: Jahresbericht 1999. Karlsruhe, 2000, S. 42-43.

Müller, Markus; Link, Norbert; Korn, Axel; Berger, Lothar; Heinze, Norbert; Krüger, Wolfgang; Schwerdtmann, Werner:

Flexible multisensor image exploitation by model based ATR.

In: Veridian ERIM International, Ann Arbor, Mich. / Infrared Information Analysis Center: Proceedings of the fourth joint international military sensing symposium (MSS), [held at The Ecole Polytechnique, Palaiseau, France, 19-22 September 2000], Vol. 45, No. 1. Ann Arbor, Mich.: Veridian ERIM International, 2000, S. 241-252.

Müller, Markus; Krüger, Wolfgang; Heinze, Norbert:

Model based target and background characterization.

In: Watkins, Wendell R. (Eds.); Society of Photo-Optical Instrumentation Engineers (SPIE), Bellingham, Wash.: Targets and backgrounds VI: Characterization, visualization and the detection process. Orlando, Florida, 24-26 April 2000. Bellingham, Wash.: SPIE, 2000, S. 88-92. (Proceedings of SPIE 4029).

Müller, Markus:

Saliency measures in cluttered IR images for ATR.

In: Watkins, Wendell R. (Eds.); Society of Photo-Optical Instrumentation Engineers (SPIE), Bellingham, Wash.: Targets and backgrounds: Characterization and representation V. Orlando, Florida, 5-7 April 1999. Bellingham, Wash. : SPIE, 1999, S. 150-154. (Proceedings of SPIE 3699).

Müller, Markus; Korn, Axel:

Automatic target detection in cluttered IR images.

In: Watkins, Wendell R. (Eds.); Society of Photo-Optical Instrumentation Engineers (SPIE), Bellingham, Wash.: Targets and backgrounds: Characterization and representation IV. Orlando, Florida, 13-15 April 1998. Bellingham, Wash.: SPIE, 1998, S. 54-57. (Proceedings of SPIE 3375).

Müller, Markus; Geisler, Jürgen:

Interactions with ATR.

In: Spearhead Exhibitions: Battlefield Systems International (BSI 98), Vol. 1: The changing battlefield. Surrey, 1998, S. 133-140.

Korn, Axel; Müller, Markus; Sung, C.-K.:

Computer-augmented detection of targets in cluttered and low-contrast backgrounds.

In: Watkins, Wendell R. (Eds.); Society of Photo-Optical Instrumentation Engineers (SPIE), Bellingham, Wash.: Targets and backgrounds: Characterization and representation III. Orlando, Florida, 21-23 April 1997. Bellingham, Wash.: SPIE, 1997, S. 278-283. (Proceedings of SPIE 3062).

12 Literaturverzeichnis

Beyerer, J., et al. 2012. *visIT - Zivile Sicherheit*. Karlsruhe : Fraunhofer IOSB, 2012.

BMBF. 2007. *Das 7. EU-Forschungsrahmenprogramm*. Berlin : Bundesministerium für Bildung und Forschung (BMBF), Referat Öffentlichkeitsarbeit, 2007.

Brüstle, S. und Heinze, N. 2012. Archiving image sequences considering associated geographical and nongeographical attributes. *Conference Proceedings SPIE*. 2012.

Bullinger, H. J. und Röthlein, B. 2012. *Morgenstadt - Wie wir morgen leben: Lösungen für das urbane Leben der Zukunft*. München : Carl Hanser Verlag, 2012. S. 159-180.

Bunke, Horst und Niemann, Heinrich. 1987. *Künstliche Intelligenz in Bild- und Sprachanalyse (Leitfäden und Monographien der Informatik)*. s.l. : Vieweg + Teubner Verlag, 1987.

Burt, Peter J. und Adelson, Edward H. 1983. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*. 1983, Bd. 4, 31, S. 532-540.

Canny, J. F. 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1986, 8, S. 679-698.

Canny, J. F. 1983. Finding Edges and Lines in Images. *MIT Artificial Intelligence Laboratory TR-720*. 1983.

Chandrakasan, A. P., et al. 1995. Optimizing power using transformations. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. 14, 1995, Bd. 1.

Collins, Robert T., et al. 2000. A System for Video Surveillance and Monitoring. *VSAM Final Report*. 2000.

- Cremers, D., et al. 2002.** Diffusion Snakes: Introducing statistical shape knowledge into the Mumford-Shah functional. *International Journal of Computer Vision*. 50, 2002, Bd. 3.
- Daniilidis, K. 1992.** *Zur Fehlerempfindlichkeit in der Ermittlung von Objektbeschreibungen und relativen Bewegungen aus monokularen Bildfolgen.* Dissertation, Fakultät für Informatik, Universität Karlsruhe (TH) : s.n., 1992.
- Esswein, Martin. 2012.** *Softwarearchitektur ABUL 4.0.* IOSB, Karlsruhe : s.n., 2012.
- EU. 2011.** *Horizon 2020 - The Framework Programme for Research and Innovation.* Brussels : European Commission, 2011.
- Flynn, M. J. 1972.** Some Computer Organizations and their Effectiveness. *IEEE Transactions on Computers*. 1972, Bde. C-21, S. 948-960.
- Geman, S. und Geman, D. 1984.** Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 6, 1984, Bd. 6.
- Gibson, J. J. 1950.** *The Perception of the Visual World.* s.l. : Houghton-Mifflin, 1950.
- Giloi, W. K. 1981.** *Rechnerarchitektur.* Berlin, Heidelberg, New York : Springer, 1981.
- Grady, L. und Alvino, C. 2008.** Reformulating and optimizing the mumford-shah functional on a graph - a faster, lower energy solution. *European Conference on Computer Vision (ECCV)*. 2008.
- Grimm, M. 2003.** *Zur Eignung invarianter Momente für die Klassifikation von Objekten in Bildern.* Diplomarbeit, Universität Karlsruhe (TH) : s.n., 2003.
- Hansen, Michael W. und Higgins, William E. 1997.** Relaxation Methods for Supervised Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 19, 1997, Bd. 9.

- Haralik, R. M., Shanugam, K. und Dinstein, I. 1973.** Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*. 1973, Bd. 3, 6.
- Hecht, E. 2009.** *Optik*. s.l. : Oldenbourg, 2009.
- Heinbach, W. 1988.** *Gehörgerechte Repräsentation von Audiosignalen durch das Teiltonzeitmuster*. s.l. : Dissertation, Technische Universität München, 1988.
- Heinze, N., Esswein, M. und Krüger, W. 2010.** Image Exploitation algorithms for reconnaissance and surveillance with UAV. *Conference Proceedings SPIE*. 2010.
- Heinze, Norbert, et al. 2008.** Automatic image exploitation system for small UAVs. 2008, Bde. Proc. SPIE 6946, Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications V, 69460G.
- Hellmann, R. 2013.** *Rechnerarchitektur*. s.l. : Oldenbourg Wissenschaftsverlag, 2013.
- Hild, J., Brüstle, S. and Heinze, N. 2013.** Gaze interaction in UAS video exploitation. *Conference Proceedings SPIE*. 2013.
- Hopfield, J.J. 1982.** Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*. 1982, Bd. 79.
- INTEL. 2007.** *INTEL SSE4 Programming Reference*. s.l. : INTEL Reference Number D91561-001, 2007.
- Ising, Ernst. 1925.** Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*. 31, 1925, Bd. 1.
- Itti, L. 2003.** Visual Attention. [Hrsg.] M. A. Arbib. *The Handbook of Brain Theory and Neural Networks*. 2, 2003, S. 1196-1201.
- Jüngling, K. and Arens, M. 2011.** Detecting and tracking people and their body parts in infrared. *Proceedings 8th IEEE International Conference on Image Processing, Computer Vision, and Pattern Recognition*. 2011.

- Kanade, T. 1978.** Region Segmentation: Signal vs. Semantics. [Hrsg.] T. Sakai. *Proceedings of the Fourth International Joint Conference on Pattern Recognition.* 1978, S. 95-105.
- Korn, A. 1985.** Combination of different spatial frequency filters for modelling edges and surfaces in gray-value pictures. *Proceedings Computer Vision & Robots SPIE.* 1985, 595.
- Korn, Axel. 1988.** Toward a Symbolic Representation of Intensity Changes in Images. *IEEE Trans. Pattern Analysis and Machine Intelligence, PAMI 10, No. 5.* 1988.
- . **1995.** *Verarbeitung und Auswertung von Signalen bildgebender Sensoren.* Düsseldorf : Fortschrittsberichte VDI Reihe, 1995. Bd. 10.
- Korn, Axel, Müller, Markus and Sung, C.-K. 1997.** Computer-Augmented Detection of Targets in Cluttered and Low-Contrast Backgrounds. *Proceedings of SPIE, Target and Background Representation.* 1997.
- Krüger, Wolfgang. 2001.** Robust and efficient map-to-image registration with line segments. 2001, Bd. Machine Vision and Applications 13.
- Lakshmi Devasena, C., Revathi, R. und Hemalatha, M. 2011.** Video Surveillance Systems - A Survey. *IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1.* July 2011.
- Lehmer, D.H. 1957.** Combinatorial problems with digital computers. *Proc. of the Fourth Canadian Math. Congress.* 1957.
- Lowe, D. 2004.** Distinctive Image Features from Scale-Invariant Keypoints. 2004, Bd. International Journal of Computer Vision.
- Magnin, L. and Alikacem, H. 1999.** GenA: Multiplatform Generic Agents. *MATA'99 First International Workshop on Mobile Agents for Telecommunication.* 1999.
- Monari, Eduardo. 2011.** *Dynamische Sensorselektion zur auftragsorientierten Objektverfolgung in Kameranetzwerken.* Dissertation, Karlsruher Institut für Technologie (KIT), Fakultät für Elektrotechnik und Informationstechnik : s.n., 2011.

- Moore, Gordon E. 1965.** Cramming more components onto integrated circuits. *Electronics*. 1965, Bd. 8, 38.
- Müller, M. 2001.** *Entnahmestrategien zur Objektvereinzelung mittels Industrierobotern*. Dissertation, Fakultät für Informatik, Universität Karlsruhe (TH) : s.n., 2001.
- Müller, Markus und Heinze, Norbert. 2013.** ABUL - Systemübersicht. Fraunhofer IOSB : s.n., 2013.
- Müller, Markus und Korn, Axel. 1999.** Automatic Target Detection in Cluttered IR Images. *Target and Background Representation*. 1999.
- Müller, Markus, et al. 1996.** WITMUS - Wissensbasierte, teilautomatische Bildauswertung für die multisensorielle Aufklärung. 3. Zwischenbericht, 1996.
- Müller, Markus, et al. 1997.** WITMUS - Wissensbasierte, teilautomatische Bildauswertung für die multisensorielle Aufklärung. 1997, Bd. 4. Zwischenbericht.
- Müller, Markus, et al. 1999.** *WITMUS - Wissensbasierte, teilautomatische Bildauswertung für die multisensorielle Aufklärung*. Karlsruhe : Fraunhofer IOSB, 1999.
- Müller, Markus, et al. 2001.** WITMUS II - Wissensbasierte, teilautomatische Bildauswertung für die multisensorielle Aufklärung. 2001, Bd. 3. Zwischenbericht.
- Müller, Markus, Krüger, Wolfgang und Saur, Günter. 2007.** Robust image registration for fusion. *Information Fusion*. 2007, Bd. 8.
- Müller, Markus, Schreer, Oliver and Monica, Saenz Lopez. 2007.** Real-time image processing and fusion for a new high-speed dual-band infrared camera. [ed.] Gerald C. Holst. *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVIII*. 2007.
- Müller, Thomas und Müller, Markus. 2011.** CART V: Recent advancements in computer aided camouflage assessment. [Hrsg.] Gerald C. Holst und Keith A. Krapels. 2011, Bde. SPIE Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXI.

- Mumford, David und Shah, Jayant. 1989.** Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems. *Communications on Pure and Applied Mathematics*. XLII, 1989, Bd. 5.
- Nagel, H.-H. 1979.** Über die Repräsentation von Wissen zur Auswertung von Bildern. [Hrsg.] J. P. Foith. *Angewandte Szenenanalyse, DAGM Symposium*. 1979, S. 3-21.
- Oueslati, Wided und Akaichi, Jalel. 2010.** A Survey on Data Warehouse Evolution. *International Journal of Database Management Systems (IJDMS)*. 2010, Bd. 2, 4.
- Panchal, P. M., Panchal, S. R. und Shah, S. K. 2013.** A Comparison of SIFT and SURF. 2013, Bde. *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 1, Issue 2.
- PERSPECTIVES. 2013.** *Smart Cities - Städte von hoher digitaler Intelligenz*. s.l. : Intergraph Security, Government & Infrastructure, 2013.
- Pock, T., et al. 2009.** A convex relaxation approach for minimizing the Potts model. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2009.
- Renier, L., et al. 2005.** The Ponzo illusion with auditory substitution of vision in sighted and early-blind subjects. *Perception*. 34, 2005.
- Schönbein, Rainer. 1989.** Effiziente Verfahrensentwicklung für die Bildauswertung durch objektorientierten, adaptierbaren Dialog und videoschnelle Verarbeitung im Funktionsverbund KIBAS-VISTA. [Hrsg.] H. Burkhardt, K. H. Höhne und B. Neumann. Berlin, Heidelberg, New York : Springer-Verlag, 1989. Bde. DAGM-Symposium Mustererkennung, 11.
- Sedgewick, R. and Wayne, K. 2011.** *Algorithms*. s.l. : Pearson Education, Inc., 2011.
- Sutter, Herb. 2005.** The free lunch is over: A fundamental turn toward concurrency in software. *Dr. Dobbs' Journal*. 30, 2005, Bd. 3.

- Teutsch, M., Krüger, W. und Heinze, N. 2011.** Detection and classification of moving objects from UAVs with optical sensors. *Conference Proceedings SPIE*. 2011.
- Treisman, A. and Gelade, G. 1980.** A Feature Integration Theory of Attention. *Cognitive Psychology*. 1980, 12, pp. 97-136.
- Treisman, A. 1986.** Features and Objects in Visual Processing. *Scientific American*. 1986, 255.
- Tsai, A, Yezzi, A. J. und Willsky, A. S. 2001.** Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Transactions on Image Processing*. 10, 2001, Bd. 8.
- Ullman, S. 1995.** *High-level Vision*. s.l. : The MIT Press Cambridge, Massachusetts London, England, 1995.
- Vagts, Hauke-Hendrik. 2013.** *Privatheit und Datenschutz in der intelligenten Überwachung: Ein datenschutzgewährendes System, entworfen nach dem "Privacy by Design" Prinzip*. Karlsruhe : Dissertation, Fakultät für Informatik, Karlsruher Institut für Technologie (KIT), 2013.
- Vese, L. und Chan, T. 2002.** A multiphase level set framework for image processing using the Mumford-Shah functional. *International Journal of Computer Vision*. 50, 2002, Bd. 3.
- Viterbi, A. 1967.** Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory* 13. 1967, Bd. 2.
- Waltz, D. 1975.** Understanding Line Drawings of Scenes with Shadows. [Hrsg.] P. H. Winston. *The Psychology of Computer Vision*. 1975.
- Wirth, Niklaus. 1995.** A Plea for Lean Software. *Computer Magazin*. IEEE Computer Society Press Los Alamitos, CA, USA, 1995, Bd. 2, 28.

Zimmermann, Georg und Kories, Ralf. 1989. Eine Familie von nichtlinearen Operatoren zur robusten Auswertung von Bildfolgen. [Hrsg.] Werner Schwerdtmann. 1989, Bd. Ausgewählte Verfahren der Mustererkennung und Bildverarbeitung.