Karlsruher Schriften
zur Anthropomatik

Band 34

Jürgen Beyerer, Alexey Pak and Miro Taphanel (Eds.)

Proceedings of the 2017 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory

KIT Scientific Publishing

Jürgen Beyerer, Alexey Pak and Miro Taphanel (Eds.)

Proceedings of the 2017 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory

Karlsruher Schriften zur Anthropomatik
Band 34
Herausgeber: Prof. Dr.-Ing. habil. Jürgen Beyerer

Eine Übersicht aller bisher in dieser Schriftenreihe
erschienenen Bände finden Sie am Ende des Buchs.

Proceedings of the 2017 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory

Edited by
Jürgen Beyerer, Alexey Pak and Miro Taphanel

KIT Scientific Publishing

Impressum

KIT Scientific Publishing

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

KIT Scientific Publishing is a registered trademark
of Karlsruhe Institute of Technology.
Reprint using the book cover is not allowed.

www.ksp.kit.edu

Print on Demand 2018 – Gedruckt auf FSC-zertifiziertem Papier

ISSN 1863-6489
ISBN 978-3-7315-0779-6
DOI 10.5445/KSP/1000081314

Preface

In 2017, the annual joint workshop of the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB) and the Vision and Fusion Laboratory (IES) of the Institute for Anthropomatics, Karlsruhe Institute of Technology (KIT) has again been hosted by the town of Triberg-Nussbach in Germany.

For a week from July, 30 to August, 5 the PhD students of the both institutions delivered extended reports on the status of their research and participated in thorough discussions on topics ranging from computer vision and optical metrology to network security and neural networks. Most results and ideas presented at the workshop are collected in this book in the form of detailed technical reports. This volume provides a comprehensive and up-to-date overview of the research program of the IES Laboratory and the Fraunhofer IOSB.

The editors thank Lars Sommer, Matthias Richter, and other organizers for their efforts resulting in a pleasant and inspiring atmosphere throughout the week. We would also like to thank the doctoral students for writing and reviewing the technical reports as well as for responding to the comments and the suggestions of their colleagues.

Prof. Dr.-Ing. habil. Jürgen Beyerer
Alexey Pak, PhD
Dr.-Ing. Miro Taphanel

Contents

Automatic Inspection Planning for Optimizing the Surface Coverage in Industrial Inspection 1
Mahsa Mohammadikaji

Wavelet Based Feature Extraction in Near Infrared Spectra for Compositional Analysis of Food 15
Julius Krause

Adaptive Measurement Method for Area Chromatic Confocal Microscopy 31
Ding Luo

Deterministic Industrial Network Communication: Fundamentals 45
Ankush Meshram

Phase Detection in Medical Context: Overview and Outlook 63
Patrick Philipp

Deep Learning based Vehicle Detection in Aerial Imagery 83
Lars Sommer

Automatic Inspection Planning for Optimizing the Surface Coverage in Industrial Inspection

Mahsa Mohammadikaji

Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT), Germany
mahsa.mohammadikaji@kit.edu

Technical Report IES-2017-01

Abstract: Optical product inspection plays an important role in today's industrial manufacturing. Therefore, design of optimized solutions corresponding to the industrial requirements are essential for efficient product quality assurance. To configure an inspection setup one requires to determine the position and orientation of the cameras and illuminations as well as the optical configurations. This problem is commonly known as inspection planning. Today's optical inspection setups are mostly being designed based on trial and error requiring a lot of engineering experience and experimental work. As the design space is high dimensional, the empirical designs typically lead to suboptimal solutions and compromise between contrary requirements. In today's industry we are missing a generic automatic method to translate the inspection requirements into optimized inspection solutions. In this report we propose an optimization framework to automatically propose optimized setup solutions, by minimizing the number of acquisitions which fulfil the inspection requirements. As an example, we consider maximizing the surface coverage for the inspection of a cylinder head in a laser triangulation setup. We characterize the design space and propose different approaches to solve the problem. We finally demonstrate the planning results which successfully cover hard to reach areas on the object.

1 Introduction

A fast, automated, and precise quality inspection process is of high importance in today's industrial production. Automated inspection first made its way into

industry with Coordinate Measuring Machines (CMMs) [WPH06], the tactile probes which scan the product in a number of key points. Although CMMs deliver precise measurements, they are being more and more replaced by optical inspection techniques due to two main reasons: the very low inspection rate and their requirement to touch the object [NJ95]. The optical inspection techniques, on the other hand, offer fast touchless scans even with lower hardware costs.

The very benefits that optical methods offer are however not without an extra cost: the inspection planning. If the configuration of CMMs needed a pre-selection of the target key points on the surface, designing an optical inspection setup which would deliver the same measurement quality is not a trivial task [CBL02]. Apart from choosing among the existing inspection techniques such as fringe projection, laser triangulation, deflectometry, interferometry, and etc. (see [BLF15]), one requires to determine the position and orientation of the cameras and illuminations as well as their optical configurations. This problem is commonly referred to as inspection planning in the literature [SRR03, TTA+95]. As the design space is high dimensional, today's optical inspection setups are mostly being designed based on a trial and error process requiring a lot of engineering experience and experimental work. In addition to the high design costs, the empirical designs typically lead to suboptimal solutions and a compromise between contrary requirements. In today's industry we are missing a generic automatic method to translate the inspection requirements for a given industrial product into optimized inspection solutions. Especially for precise inspection of geometrically complex products, such as an engine block in figure 1.1, the need for an automatic inspection planning is more evident.

In this report we try to address this question by proposing an optimization method for planning the inspection of a cylinder head in a laser triangulation setup. The

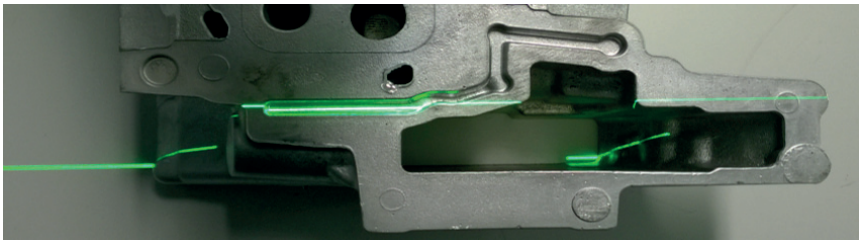


Figure 1.1: Cylinder head illuminated with a laser line

input to this optimization framework is the CAD model of the product with the desired region of interest and the inspection requirements for each region. For the automatic setup planning, we associate each design configuration to a fitness value which enables us to optimize the designed cost function to come up with optimal solutions according to the industry requirements.

The content of this report is organized as follows. In section 2 we formally define the problem of inspection planning as an optimization problem. Section 3 discusses the parameter space of a typical laser triangulation setup. In section 4, we further search for the solution of the optimization problem using two different approaches, the greedy and the combinational approach, with each of them using the Particle Swarm Optimization [BK07] method for the global optimization. In section 5 the achieved planning results will be discussed.

2 Inspection Planning Problem

Let us assume the parameter vector $\mathbf{c}$ determines all the parameters to specify a particular measurement constellation. This means parameter $\mathbf{c}$ includes all the information regarding the positioning and optical configuration of the setup. Similarly, we can define a sequence of k measurements $\{\mathbf{c}_1, ..., \mathbf{c}_k\}$. Let's further assume there is a cost function $f(\mathbf{c}_1, ..., \mathbf{c}_j) \in \mathbb{R}$ which associates each sequence of measurements to a fitness value corresponding to the quality of the measurement with a maximum value of f_{opt}. In a real measurement setup there are of course always constraints. Therefore, not every sequence of measurement can be realizable. Let's quantify all the constraints as a function $r(\mathbf{c}_1, ..., \mathbf{c}_j)$ which takes positive values if the constraints are met.

With the above definitions, the planning problem can be defined as finding the minimum number of N measurements $\{\mathbf{c}_1^*, ..., \mathbf{c}_N^*\}$ which optimize the cost function and at the same time meet the constraints. Therefore

$$\begin{aligned} & f(\mathbf{c}_1^*, ..., \mathbf{c}_N^*) = f_{opt}, \\ \text{s.t.} \quad & r(\mathbf{c}_1^*, ..., \mathbf{c}_N^*) > 0. \end{aligned}$$

For dimensional inspection, the cost function f usually encompasses either of the surface coverage, the scan resolution, or the measurement uncertainty. In our previous works [MBI+17, MBI+16] we have studied different useful cost functions for this problem.

One can easily verify that the optimization problem defined above is at least as hard as the "Set Cover" problem which is known to be NP-hard [Cor07]. The planning problem would be exactly equivalent to the set cover problem when we only consider a discrete set of constellations C for possible solutions, and the surface coverage as the cost function with no further constraints. In this scenario each constellation can cover a subset of the surface. Let us assume the union of all measurable areas by any of the constellations in C is the total area A. In the planning problem, we look for the minimum number of constellations which cover A. In the general case however, the problem is more complex. Often the number of valid constellations are not finite and the cost function is based on complex quantities such as the measurement uncertainty. The optimum value f_{opt} is also often unknown.

For large and complex object surfaces, one is often not able to search the whole parameter space and therefore, needs to resort to approximations. For the set cover problem there is a simple greedy approximation which is actually shown to be the best possible polynomial approximation to the problem [LY94]. In the context of inspection planning, the idea of the greedy approximation is to choose a constellation which covers the most of the surface (or achieves the best fitness values) and continue choosing next best constellations in the same manner. Later in section 5 we discuss the results of the greedy and non-greedy approach.

2.1 Simulations

Typical optimizations require evaluation of many different constellations which are of course not possible to be evaluated in a real setup. Therefore inspection planning always relies on simulations. The planning cost function is consequently based on evaluating the simulation results which is in general a multi-modal non-derivable function.

Evaluation of different cost functions may require different levels of realism in the simulations. The coverage of the surface can be, for instance, estimated by means of fast rasterization-based [AMHH08] simulations (milliseconds per frame), whereas evaluation of the resulting measurement uncertainty requires physically correct image formation simulations using advanced techniques [DBB16]. In such simulations, the emitted photons from the light sources are traced as they get differently scattered by the objects in the scene, all the way

to reaching the camera sensor and being converted to intensity values. In our previous work [IBM+17] we have discussed and compared the results of different simulation techniques for simulation of the images of a laser triangulation setup.

3 Design Space

In this section we discuss the degrees of freedom for the planning of a laser triangulation inspection. In such an inspection setup, we have at least one camera and one laser line projector which illuminates a profile on the object (see figure 1.1). The camera captures images of the illuminated profile which are further processed for obtaining the 3D information [MBI+16]. To scan the whole surface, the laser and camera can follow an arbitrary trajectory around the object and capture image frames all along the trajectories. Figure 3.1 depicts the cylinder head CAD model in an arbitrary constellation. For now we assume that the optical parameters of the laser and the camera (e.g. laser power, shutter time, objective f-number, etc.) are already set and are not supposed to be optimized. We solely consider the determination of the geometrical degrees of freedom of the setup. Positioning the camera and the laser as two rigid bodies has a total of 12 degrees of freedom, which can arbitrarily change along the trajectories for each single image frame. To cover a typical product with an average area comparable with the cylinder head, one needs thousands of frames. Obviously this general parametrization leads to an enormous optimization complexity.

One can however make meaningful simplifications to reduce the complexity. For instance, we can assume the scan trajectory to be a linear motion along a particular axis but allow the object instead to freely position under the sensor. The effect of laser distance to the surface can be neglected as lasers can be later focused to any particular distance. We may also parametrize the space in a way that the laser and camera both look towards a common point so that the laser always remains in the camera field of view.

Figure 3.2 illustrates the proposed 9-dimensional degrees of freedom for parametrization of one acquisition. Each acquisition is defined as completely scanning the object along the predefined scan direction. Therefore, instead of planning for each single frame we plan for a number of N acquisitions. In the proposed design space, we dedicate four parameters to sensor placement

(ϕ, θ, τ, d) and five to the object $(\alpha, \beta, \gamma, \Delta x, \Delta z)$. Similar to spherical coordinates, camera placement is determined using polar angle θ, azimuthal angle ϕ, and distance d. The laser holds a triangulation angle τ to the camera, with positive values corresponding to bright field illumination and negative values for dark field illumination. The distance of the laser to the origin is set to a predefined value and does not change during the optimization. The rotation angle ϕ rotates both the camera and the laser to always keep the laser line aligned with the image rows. The object can be freely positioned under the sensor by a 3D rotation using rotation angles α, β, γ, as well as a translational vector. The translation has however only two degrees of freedom because during the acquisition the object is completely scanned in one direction and the translational components parallel to the scan direction do not introduce any new constellations. In the proposed parametrization we consider the y-axis as the scan direction and consider it invariable. Due to rotational degrees of freedom of the sensor and the object, variations of the scan direction within the xy plane will be redundant and do not count as an extra degree of freedom.

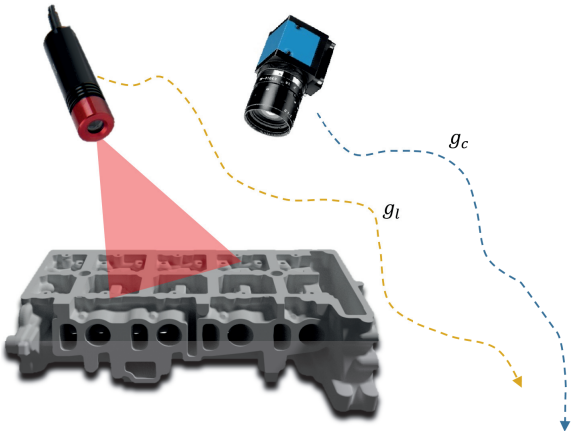


Figure 3.1: Camera and laser can move along arbitrary trajectories g_c and g_l to scan the whole surface.

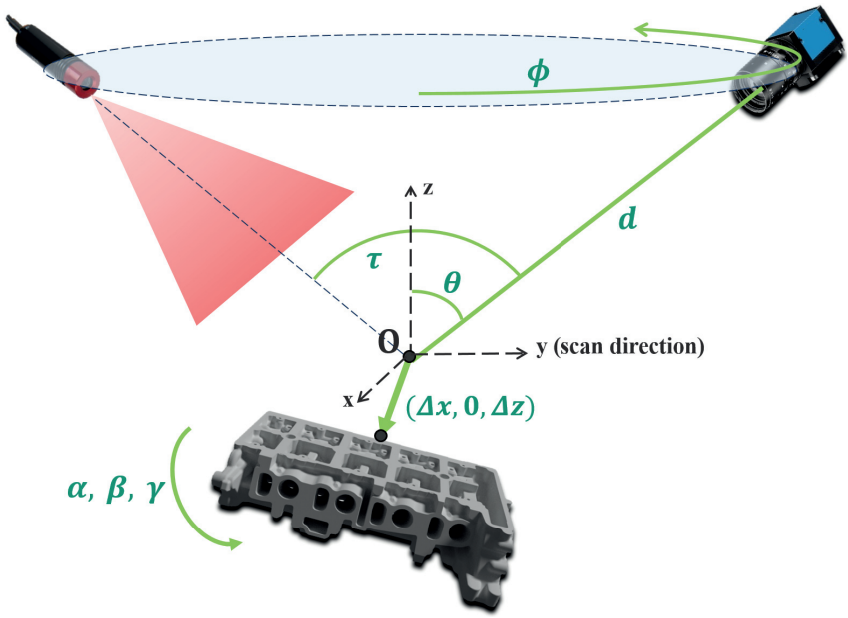


Figure 3.2: Proposed 9D design space for one acquisition

4 Optimization

After definition of the degrees of freedom, in this section we go back to the main optimization problem in equation 2. The main difficulty with the optimization is that we not only want to optimize the constellation parameters but also minimize the number of acquisitions. For a fixed number of M acquisitions, one has an optimization problem with $9\mathrm{M}$ parameters for which we can use standard optimization algorithms. To find the optimal value for M one needs another optimization on possible values of M. The psudo-code in algorithm 4.1 can be a potential solution to the problem.

Algorithm 4.1 Inspection planning algorithm - combinational

Result: Minimum number of acquisition M^*, constellation parameters $\mathbf{p}^*$
$\mathrm{M} = 0$, $\mathbf{p} = [\,]$
$\mathrm{lastFitness} = 0$, $\mathrm{nextFitness} = 0$
repeat
    `/* try the solution to M+1 acquisitions */`
    $\mathrm{M} = \mathrm{M} + 1$
    $\mathrm{lastFitness} = \mathrm{nextFitness}$
    $[\mathrm{nextFitness}, \mathbf{p}] = optimizeMultipleAcquisitions(\mathrm{M})$
until $(\mathrm{nextFitness} - \mathrm{lastFitness} \leq \mathrm{T}$ or $\mathrm{surfaceIsCovered})$;
$\mathrm{M}^* = \mathrm{M}$, $\mathbf{p}^* = \mathbf{p}$

In the above algorithm, one goes through a $9(\mathrm{M}+1)$-dimensional optimization in each iteration to obtain the optimized fitness value and the corresponding parameter vector $\mathbf{p}$, and compares the gain of having $\mathrm{M}+1$ optimized acquisitions compared to M optimized acquisitions. If the gain of allowing one more acquisition is zero or insignificant we stop the optimization and return the result of the previous step. Of course one can modify the linear search in this algorithm into a binary search to significantly increase the performance; however, even for moderate values of M like 20, the high-dimensional optimization ($9 \times 21 = 189$ free parameters) imposes a great overhead. This is not only a problem from the performance point of view (the algorithm is in the complexity class NP), but it makes it significantly less probable for a heuristic optimization algorithm to find the global optimum of a non-convex non-derivable high-dimensional cost function.

As mentioned earlier, the best polynomial approximation to the set cover problem is the greedy approach. In this algorithm, one already uses the previously optimized constellations and adds only a single constellation to the previous ones. This keeps the optimization overhead constant with the increase of the acquisitions; however, this may result in more acquisitions than the theoretical optimal value. For the set cover problem, the approximation factor is shown to be bounded by a logarithmic upper-bound [LY94]. Algorithm 4.2 contains the pseudo-code of the greedy approximation to the inspection planning algorithm.

Algorithm 4.2 Inspection planning algorithm - greedy approximation

Result: Minimum number of scans M^* with parameters $\mathbf{p}^*$
$\mathrm{M} = 0$, $\mathbf{p} = [\,]$
$\mathrm{lastFitness} = 0$, $\mathrm{nextFitness} = 0$
repeat
  $\mathrm{lastFitness} = \mathrm{nextFitness}$
  `/* 9-dimensional optimization */`
  $[\mathrm{nextFitness}, \mathbf{c}] = findNextBestScan(\mathbf{p})$
  `/* append to previous scans */`
  $append(\mathbf{p}, \mathbf{c})$
  $\mathrm{M} = \mathrm{M} + 1$
until $(\mathrm{nextFitness} - \mathrm{lastFitness} \leq \mathrm{T}$ or surfaceIsCovered$)$;
$\mathrm{M}^* = \mathrm{M}$, $\mathbf{p}^* = \mathbf{p}$

4.1 Particle Swarm Optimization

Using either of the algorithms for solving the inspection planning problem, we need a global optimizer. As mentioned earlier, the cost function is based on evaluating simulations, taking any complex form with multiple local optimums. Therefore, derivative-free randomized heuristic search algorithms seem to be the right choice for the global optimization problem.

Among the many existing global optimizaters in the literature (see [Wei09] for a good review), the Particle Swarm Optimization (PSO) has been widely applied in many fields. PSO is a variant of swarm-based intelligent methods [BDT99] whose main idea is to imitate the behavior of biological species which live in colonies such as birds, ants, and bees. We observe that these species exhibit a high intelligence in their social activities such as searching for food, although they exhibit simple individual acts. The PSO algorithm is especially inspired by the way birds communicate when searching a field for food [Wei09]. The approach of PSO is very intuitive. For optimizing a cost function $f(\mathbf{x})$, one initiates a number of K random particles in the parameter space (the swarm size). The particles are not only individual searchers in the paramter space (similar to

many parallel simulated-annealing [Wei09] optimizers), but they also communicate with each other and share the results of their local searches which guides them for the rest of the search.

Every particle in this algorithm is composed of three vectors: its current position in the search space $\mathbf{x}_i$, its best individually found position so far $\mathbf{b}_i$, and its velocity $\mathbf{v}_i$. The particles are also aware of the best position $\mathbf{g}_i$ found by the rest of the particles so far. This might be actually the globally best found position, or the best position that the particle has so far heard from those who have communicated with it. The algorithm updates the particles at each iteration by updating their velocities and positions according to

$$\begin{aligned}\mathbf{v_i}^{t+1} &= \mathbf{v_i}^{t} + \epsilon_1(\mathbf{b}_i^t - \mathbf{x}_i^t) + \epsilon_2(\mathbf{g}_i^t - \mathbf{x}_i^t),\\ \mathbf{x}_i^{t+1} &= \mathbf{x}_i^t + \mathbf{v_i}^{t}.\end{aligned}$$

The random factors ϵ_1 and ϵ_2 weight tendencies of the particle to search further towards the local and global optimum. A particle also has a tendency to keep its previous direction, therefore we also add the $\mathbf{v_i}^t$ term to the velocity update rule. The position is simply computed as moving from the previous position along the updated velocity.

Since the introduction of PSO, there have been a few standardizations proposed which give recommendations on choosing the parameters of the algorithm like the swarm size, the communication structure, and weightings. In this work, we have orientated the PSO implementation based on the standardization given in 2007 [BK].

5 Coverage Planning Results

The cylinder head object contains hard to reach areas such as deep intake and exhaust manifolds. Therefore, maximizing the surface coverage with minimized number of scans is a non-trivial problem which is of high interest for the industry. In this section, we present the inspection planning results with the goal of maximizing the surface coverage. For evaluating the surface coverage, we use a fine mesh model of the object as shown in figure 5.1. The granularity of the mesh elements must correspond to the required inspection resolution. The surface coverage can be then evaluated as the area of all the patches which have been measured with at least one point.

For the optimization, we have used the proposed parameter space in figure 3.2 with a simplification of setting the camera distance d to a constant value of 0.5 meter. This choice can be justified by the fact that the object translation component along the z-axis introduces very similar effects to changing the camera distance. However, in the future we intend to get the results with the full degrees of freedom. As optimization constraints, we assume valid ranges for each of the parameters, based on the degrees of freedom of the real physical setup. For the current results we have bounded the parameters to

$$\phi \in [0°, 360°), \quad \theta \in [0°, 80°], \quad \tau \in [-80°, -10] \cup [10°, 80],$$
$$\alpha, \beta, \gamma \in [0°, 360°), \quad \Delta x \in [-0.4, 0.4]\, m, \quad \Delta z \in [-0.5, 0.5]\, m.$$

Many parameters can already take their full range, such as ϕ. Others such as the triangulation angle τ must be constrained to deliver meaningful measurements.

Figure 5.2 compares the results of the greedy vs. the combinational approach. This chart displays the optimized measurable area for each number of acquisitions. The orange line determines the full area of the object, which is however not fully measurable because some areas are either completely unreachable or require constellations which violate the optimization constraints. The blue dotted graph displays the improvements of the greedy surface coverage planning vs. the number of acquisitions. One can see that the algorithm makes rather big improvements at the beginning; however, the contribution of the next acquisitions gradually reduces until it falls below the threshold for the 30th acquisition. For comparison, we have also applied the combinational planning algorithm for

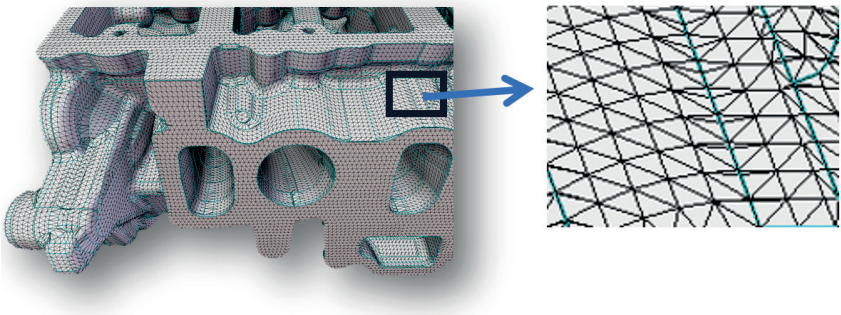


Figure 5.1: Cylinder head fine mesh model for evaluating surface coverage

5, 10, 15, and 20 acquisitions, where we allow the optimizer to combinationally optimize all the acquisitions together. In theory the combinational approach is able to find the global optimum as it looks for every possible combination of constellations. The greedy approach however, might be trapped in local optimum solutions as the constellations are optimized in a non-reversible approach. In practice, however, we see that the results of the combinational approach can even underperform the greedy method. This is due to the fact that it is less probable for a heuristic random optimizer to find the optimum of a high dimensional problem, compared to a problem with significantly less degrees of freedom.

Figure 5.3 illustrates the resulting point cloud of the object after applying the 30 optimized acquisitions obtained by the greedy planning. As it can be seen, the deep cavities corresponding to the intake and exhaust manifolds have been covered.

6 Summary and Future work

In this report we proposed and implemented an optimization framework for automatic optimization of industrial inspection setups. We went through the

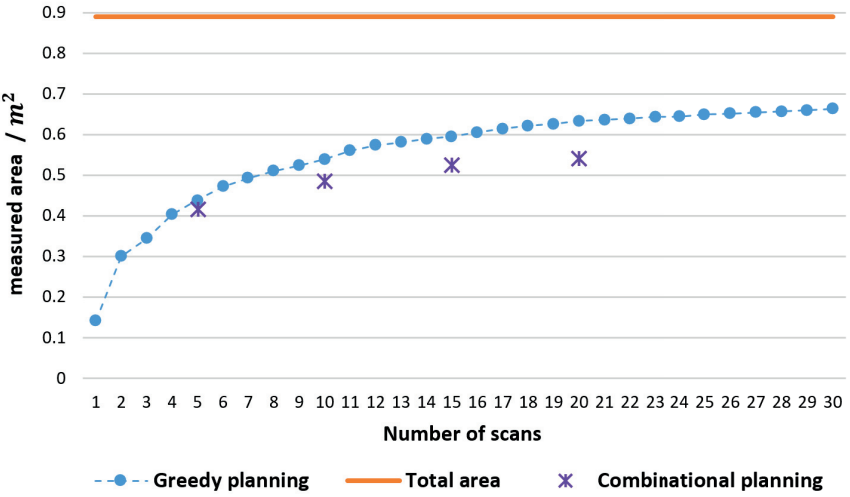


Figure 5.2: Optimized surface coverage vs. number of acquisitions

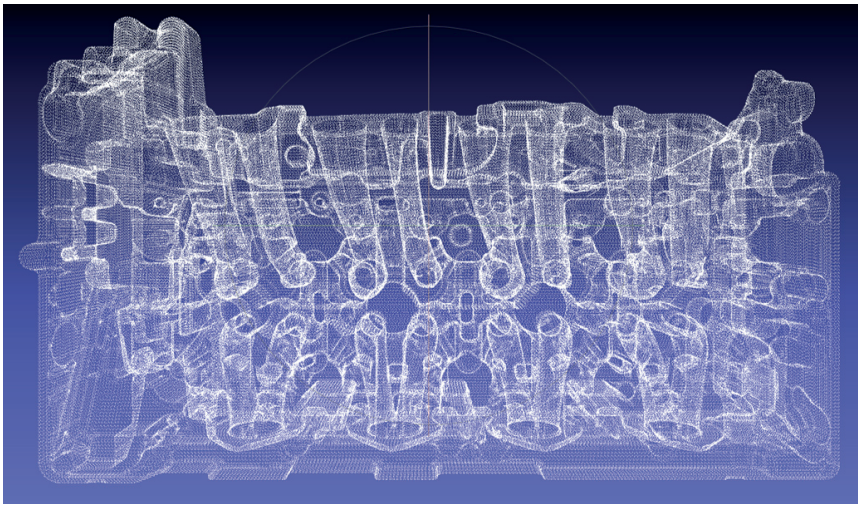


Figure 5.3: Point cloud of the resulting measurement after applying 30 optimized acquisitions obtained by the greedy planning.

parametrization of the design space for inspecting a cylinder head in a laser triangulation setup and compared different approaches for finding the minimized number of acquisitions which maximize the surface coverage.

It is very insightful to calculate the actual measurable area on the cylinder head to better evaluate the achieved coverage results. As a future work, we plan to apply methods for calculating ambient occlusion [Mil94] to calculate the area measurable for a given set of inspection constraints. In addition, there is also potentials to combine the greedy and the combinational optimization approaches to benefit from both. In a hybrid approach, one can use the greedy approach to come up with good starting points with less optimization overhead and further apply the combinational approach on the achieved suboptimal results to globally improve the results.

Bibliography

[AMHH08] Tomas Akenine-Möller, Eric Haines, and Naty Hoffman. *Real-Time Rendering*. A. K. Peters, Ltd, Natick, MA, USA, 3 edition, 2008.

[BDT99] Eric Bonabeau, Marco Dorigo, and Guy Theraulaz. *Swarm intelligence: from natural to artificial systems*. Oxford University Press, 1999.

[BK07] Daniel Bratton and James Kennedy. Defining a standard for particle swarm optimization. In *Proc. 2007 IEEE Swarm Intelligence Symposium*, pages 120–127, 2007.

[BLF15] Jürgen Beyerer, Fernando Puente León, and Christian Frese. *Machine Vision: Automated Visual Inspection: Theory, Practice and Applications*. Springer, 2015.

[CBL02] A. Contri, P. Bourdet, and C. Lartigue. Quality of 3D digitised points obtained with non-contact optical sensors. *CIRP Annals - Manufacturing Technology*, 51(1):443–446, 2002.

[Cor07] Thomas H. Cormen. *Introduction to algorithms*. MIT Press, Cambridge, Mass., 2 edition, 2007.

[DBB16] Philip Dutre, Philippe Bekaert, and Kavita Bala. *Advanced global illumination*. CRC Press, 2016.

[IBM+17] S. Irgenfried, S. Bergmann, M. Mohammadikaji, J. Beyerer, C. Dachsbacher, and H. Wörn. Image formation simulation for computer-aided inspection planning of machine vision systems. In Jürgen Beyerer and Fernando Puente León, editors, *Proc. SPIE Optical Metrology*, SPIE Proceedings, page 1033406. SPIE, 2017.

[LY94] Carsten Lund and Mihalis Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41(5):960–981, 1994.

[MBI+16] M. Mohammadikaji, S. Bergmann, S. Irgenfried, J. Beyerer, C. Dachsbacher, and H. Wörn. A framework for uncertainty propagation in 3D shape measurement using laser triangulation. In *IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 6–11, 2016.

[MBI+17] M. Mohammadikaji, S. Bergmann, S. Irgenfried, J. Beyerer, C. Dachsbacher, and H. Wörn. Probabilistic surface inference for industrial inspection planning. In *Proc. 2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1008–1016, 2017.

[Mil94] Gavin Miller. Efficient algorithms for local and global accessibility shading. In Dino Schweitzer, Andrew Glassner, and Mike Keeler, editors, *Proc. SIGGRAPH 1994*, pages 319–326, 1994.

[NJ95] Timothy S. Newman and Anil K. Jain. A survey of automated visual inspection. *Computer Vision and Image Understanding*, 61(2):231–262, 1995.

[SRR03] William R. Scott, Gerhard Roth, and Jean-François Rivest. View planning for automated three-dimensional object reconstruction and inspection. *ACM Computing Surveys*, 35(1):64–96, 2003.

[TTA+95] Konstantinos Tarabanis, Roger Y Tsai, Peter K Allen, et al. The MVP sensor planning system for robotic vision tasks. *IEEE Transactions on Robotics and Automation*, 11(1):72–85, 1995.

[Wei09] Thomas Weise. Global optimization algorithms – theory and application. *Self-Published*, 2009.

[WPH06] A Weckenmann, G Peggs, and J Hoffmann. Probing systems for dimensional micro-and nano-metrology. *Measurement Science and Technology*, 17(3):1–504, 2006.

Wavelet Based Feature Extraction in Near Infrared Spectra for Compositional Analysis of Food

Julius Krause

Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT), Germany
julius.krause@kit.edu

Technical Report IES-2017-02

Abstract: Near infrared spectroscopy is a common method for analysis of food, soil and pharmaceutical products. New developments in sensor technology, like hyperspectral camera systems and mobile spectrometers, allow broad applications of spectroscopy with devices out of specialized laboratories. Therefore, it is necessary to develop robust algorithms for classification and regression, regardless of the device. The key to robust analysis lies in data preparation to get standardized spectral information from each device. Wavelet based feature extraction could be a possible method to compress spectral data to its material specific absorption information. A method for wavelet based feature extraction, which also reduces the influence from elastic scattering effects is proposed in this report.

1 Introduction

In order to ensure the high standards of food quality, monitoring measurements are required throughout the entire production process right up to the customer. Optical spectroscopy in the visible and near-infrared spectrum can be used as a non-destructive and non-contact measuring method on foods for quality determination. Compared to laboratory tests, the result of an optical measurement is immediately available [LGGFR17].

In the future, the development of compact and cost-effective sensor technology will facilitate the dissemination of spectroscopy. Due to advancing developments

in microsystems technology, it has been possible to integrate different measurement methods like tunable Fabry-Perot filter, fourier-transform or scanning grating systems into miniaturized sensors. A series production at prices of a few US dollars has already been announced. A ”food scanner” is just one possible application. The integration of these sensors in the Internet of Things (IoT) or a smartphone is also possible and opens up a variety of other applications in the field of quality and process control [RDC17, DWKR16].

The comparability of spectroscopic data across different devices is one of the great challenges in spectroscopy. It is gaining in importance as networking of spectral sensors grows. In the history of spectroscopy, many approaches to spectral preprocessing [RvdBE09] and transfer of models [FWT+02] have been developed. Wavelet transformation has also been used for pre-processing [MNHG96].

The approach presented below attempts to extract physical features from spectral data, which only represent the absorption by molecule vibrations or electron excitations. Therefore, a wavelet transformation of the spectral data with an approximation function like Gaussian or Lorentzian shape is used to get these features in connection with derivative pre-processing.

2 Physical model of the interaction between light and matter

The method presented in this article is based on the idea of describing the measurement signal by a physically motivated model. The analysis, based on physical model parameters, provides a level of abstraction in which specific disturbing influences can be specifically suppressed. At the same time, sensor-independent chemometric modeling is possible. The following section summarizes the physical factors that the author considers relevant to the theoretical signal model.

Interaction with matter leads to an extinction Q of light intensity. The extinction process depends on the wavelength and contains an overlay of different signals from different origin. In the following, the name spectral signature is used as an umbrella term for the raw signal, which is an overlay signal of chemical and

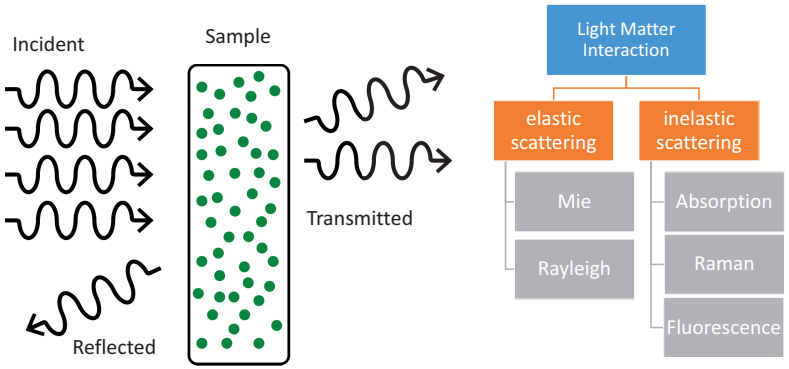


Figure 2.1: The incident light interacts with a sample. By elastic scattering within the sample, the photons are deflected by an angle θ. The most common interaction processes of the photons with the sample are shown in a block diagram.

physical properties as well as the environment. In the following, influencing variables are exemplified and summarized in three categories.

- **Chemical properties** of the sample, which are determined by absorption and fluorescence effects, which are related to specific excitations of electron states and molecular vibrations and thus produce a material-specific spectral signature.
- **Physical properties**, which depend, for example, on the shape of the sample, a surface condition of the sample or, in the case of a sample in powder form, on a degree of grinding of the powder and other properties. There is also an influence of the geometry between the measurement object and the sensor or the light sources and the sensor. This influence changes the measured spectral signature and thus also makes the comparability of different measuring devices more difficult.
- **Sensor properties**, which are caused for example by different sensitivity or different spectral measuring ranges, complicate the comparability of measurement results of different measuring devices.

It is reasonable to assume that the three categories are independent. The scattering theory will be used, to describe these processes in detail.

The occurrence of fluorescence effects should not initially be considered in the following model. However, the analysis model shown below can be extended at any time by fluorescence effects. Inelastic scattering signals from Stokes and Antistokes processes, also known as the Raman effect, are neglected due to the signal strength from 10^{-6} to 10^{-9} compared to the output signal.

The following model summarizes the influencing variables of the measurement signal: An optical sensor detects the light emitted by a sample. The measured reflection or transmission signal is considered with respect to the light emitted from the light source. Only a part of the light emitted by the sample can be detected in the solid angle $\Delta\Omega_{\text{sensor}}$ of the detector. This extinction Q_{ext} is composed of the scattering of the incident light Q_{sca} into the solid angle Ω not detected by the sensor (Fig. 2.1). The absorption process by the electrons and the molecular states is considered independent of the scatter and added as an additional term Q_{abs}:

$$Q_{\text{ext}}(\lambda, \theta, r) \coloneqq Q_{\text{sca}}\left(\frac{r}{\lambda}, \theta\right) + Q_{\text{abs}}(\lambda)\,.$$

The model includes the particle size or micro structure with a radius r, the angle θ between the light source and sensor, and the wavelength λ dependency. The two terms of the equation are described in detail below.

2.1 Elastic Scattering Theory

Many optical systems can be well described by geometric optics. In cases where the object radius r is in or below wavelength ranges, the phenomena occurring can be well described by the scattering theory of Rayleigh and Mie [CDL02].

The Rayleigh theory can be applied to describe the light scattering by particles with radius $r < 1/10\lambda$. In this regime, the particle act as an oscillating dipole driven by the electromagnetic field. A microscopic dipole absorbs a photon in a virutal state, and the subsequent emission has the characteristic of a dipole antenna. The intensity distribution

$$I(\theta, \lambda) \propto 1/\lambda^4(1 + \cos^2\theta)$$

of the scattered light results from the probability of the individual scattering angles.

In particles and structures whose dimension correspond to the wavelength λ, plasmon resonances can be excited. A complete analytic solution of the Maxwell equations exists only for spherical objects and is described in detail in the Mie theory [Mie08]. The Mie theory can also be used for particle size determination by laser diffraction and, in particular for small objects, provides a better result than Fraunhofer diffraction. For the following analysis, however, it is sufficient to know that the solution of the Mie theory is given by the so-called Bessel functions, which are smooth and differentiable.

In summary, the elastic scattering gives a smooth and differentiable low-frequency signal contribution in the detected spectral signature.

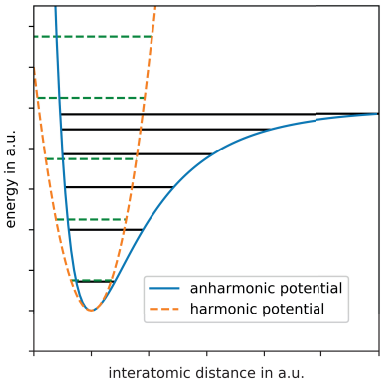
2.2 Absorption

The model of a harmonic oscillator (Lorentz oscillator) can approximately describe the absorption by molecular vibrations. Wherein the solution for determining refractive indices and absorption coefficients is reproduced substantially correctly. The dependence of wavelength of the absorption is given by the so-called dielectric function which be approximated by a Lorentz profile. In addition, the Lorentz profile corresponds to the so-called natural linewidth, which e.g. can be derived from the Fourier transform of a damped harmonic oscillator.

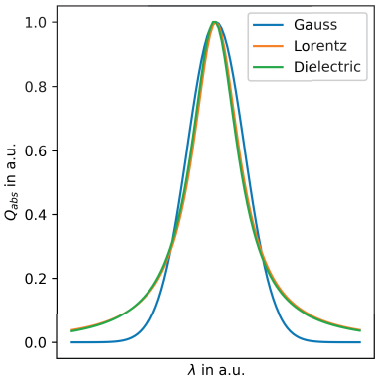
Regardless of which model is used, the exact absorption spectrum can not be calculated as long as the individual coefficients of the electric field distribution in the solid state, the anharmonicity of the molecular vibration, and the interaction with neighboring molecules are unknown. However, the course of the dependence of wavelength of a single absorbance (or emission by fluorescence) is represented approximately correctly by a bell shaped curve such as Lorentz profile or Gaussian function [Dem10].

Following the preceding qualitative analysis of the dependence of wavelength of the absorption, the relationship between an amount of substance and its absorption is now to be determined in a simple model. From the exponential attenuation of a light beam after entering a medium, the Beer-Lambert law can be derived:

$$\ln\left(\frac{I_1}{I_0}\right) = -c\eta\delta\,.$$



(a) The Lennard-Jones potential describes the interaction potential of a diatomic molecule and can be described in quadratic approximation by a harmonic oscillator. With increasing energy above the ground state, anharmonic corrections must be considered.



(b) Clearly recognizable is the good approximation of the dielectric function ϵ by a Lorentz profile. A Gaussian profile differs more, but this is justifiable because absorption lines are usually extended by disturbing effects.

Figure 2.2: Absorption on the model of diatomic molecular vibrations.

The extinction $Q_{\text{abs}} = I_1/I_0$ causes the scaling of the bell shaped absorption curve. A quantitative content determination appears possible due to the connection to the substance-dependent attenuation factor η, the concentration c, and the optical path length δ.

In summary, a single absorption can be described approximately by three parameters of a bell curve like a Gaussian.

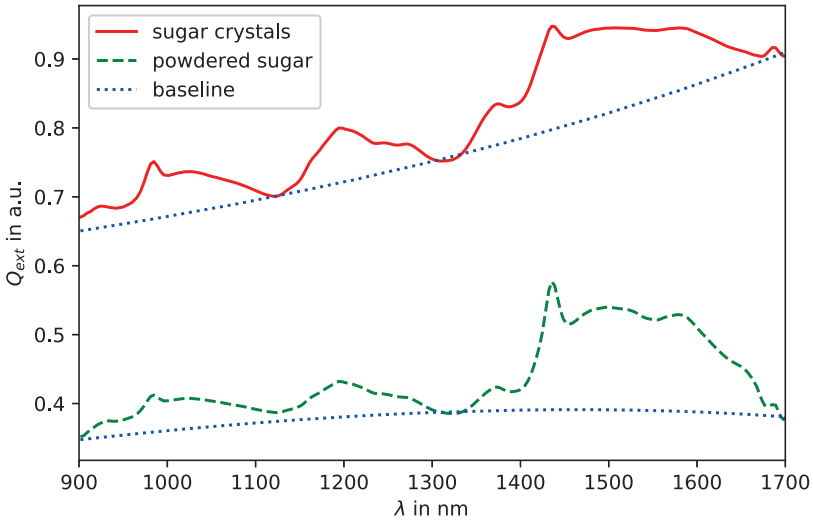


Figure 2.3: The extinction Q_{ext} of the chemically identical substances differs in scale and in the baseline due to different scattering properties and due to different particle size.

3 Wavelet Based Feature Extraction

The extinction Q_{ext} of the measurement signal in Figure 2.3 is compared to a white reflection standard. Both spectra describe the identical substance which is sugar and only the particle size differs. It can clearly be seen that the figure shows differences in the baseline of the two spectra and the signal strength. Therefore, pre-processing is needed for chemical component analysis.

The most common methods for spectral pre-treatment like scatter correction, normalization, and dimensional reduction are based on the spectral signature of a single measurement system as a whole, e.g. by inclusion of the mean signal. In addition, linear operators are destroying the connection to Beer-Lambert law. Therefore, analysis models based on these methods can not readily be used in another measurement environment.

Another established possibility for the correction of multiplicative influences is the gradient formation over the spectrum. However, the noise is amplified and human interpretation is difficult. The wavelet analysis based on the derived spectrum is intended to counteract these two disadvantages.

The wavelet analysis includes the neighbourhood information in the spectrum which counteracts the noise. The easily interpretable parameters of the approximated bell curve can be taken from the wavelet scalogram afterwards. In addition, the mean value of the wavelet transformation corrects another term of the baseline. Moreover, a later normalization based only on the absorption bands used in the model is more robust to changes in the spectral signature.

The algorithm is based on two assumptions that were explained previously:

- The baseline of the spectrum is due to the anisotropy of the elastic scattering and can be approximated by a smooth polynomial function.
- The absorption can be approximated by a bell-shaped absorption function with three parameters.

For the mathematical description, the spectrum is referred to as a function $g(\lambda)$ and the continuously differentiable bell-shaped approximation function is referred to as $\psi_{\lambda_0,s}(\lambda)$. Where λ_0 is the center and s is the width of the approximation function.

Step 1: Baseline Correction and Peak Deconvolution

The n-fold derivative of the spectrum reduces the polynomial order of the baseline, at the same time superposed peaks are unfolded [NW84]. Noise is greatly amplified by the derivative, which is why smoothing according to Savitzky-Golay is used in many cases [SG64].

In the following the fact is used that the derivative operator can also be applied to the function ψ. In the case of the second derivative, one obtains the *Mexican Hat* function, which is widely used in signal processing.

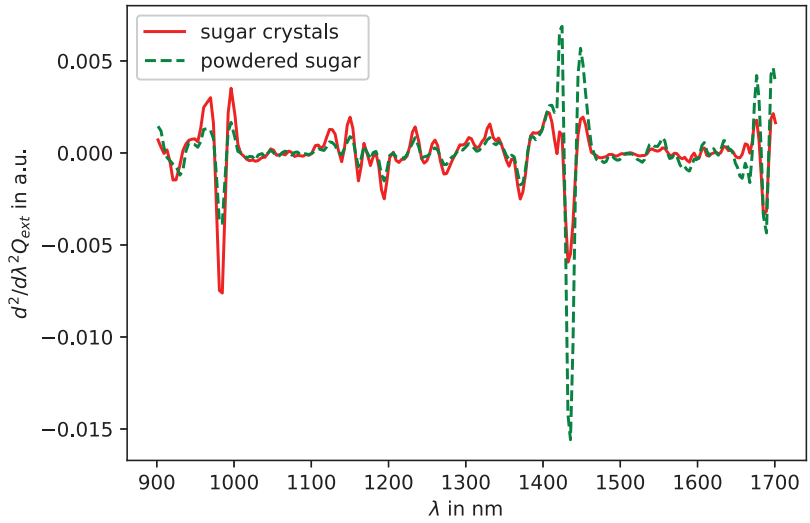


Figure 3.1: By applying the second derivative, the baseline of the spectrum was eliminated. Characteristic features are clearly shown in the shape of an inverted Mexican Hat with identical width and position in both spectra.

Step 2: Wavelet Transformation

The wavelet transformation has the character of a correlation analysis [Mal89, Mor83]. The description of the wavelet transformation as

$$\Gamma_\psi(\lambda_0, s) \coloneqq \langle \psi_{\lambda_0,s}(\lambda), g(\lambda) \rangle$$

shows this fact.

The scalar product is performed for different values of λ_0 and s, the resulting wavelet coefficients from the example of sugar is shown in a scalogram (Fig. 3.2).

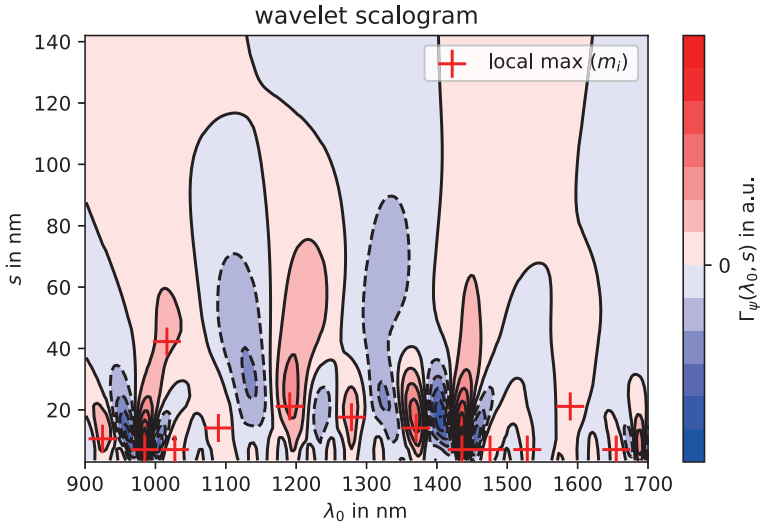


Figure 3.2: For the example of sugar, the wavelet coefficient is plotted for a selected region of s_i and λ_i. Where of the red (blue) color indicates positive (negative) energy of the wavelet coefficient. Local maxima are highlighted by red crosses.

Step 3: Feature Extraction

Local maxima of the wavelet coefficient show the location of the best match between the correlation function and the spectrum. From the coordinates of the local maxima, the parameters of each feature

$$m_i = (\lambda_i, s_i, \Gamma_\psi(\lambda_i, s_i))$$

can be found. Wherein the wavelet coefficient also indicates the height or the strength of the peak, which is linked via the Beer-Lambert law with the amount of existing ingredients. Although the amount $i \in \mathbb{N}^+$ of found features is initially not limited.

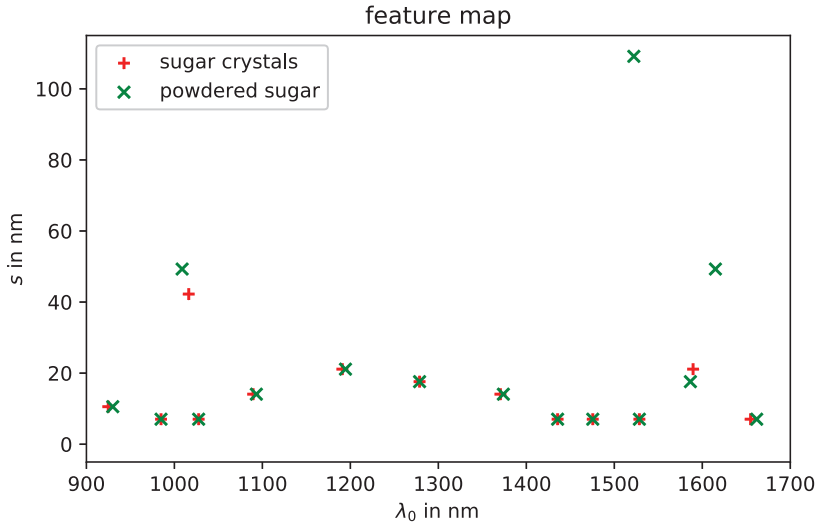


Figure 4.1: In many cases, the algorithm has extracted the identical values even for superimposed peaks.

4 Experimental Results

Using different sugar granules as an example, it could be shown that, despite different scattering properties, nearly identical features are found for the individual features (Fig. 4.1). This confirms the invariance of the found features against changes in the scattering properties.

Another example is to show that the features contain additional information of individual ingredients. For this purpose, the data set of a competition for the determination of protein in cereals by NIR spectroscopy was selected. The training dataset comprises 1488 spectra from 248 different samples, measured with 6 spectrometers, three spectrometers of the same model from two different manufacturers.

Figure 4.2 shows an example of two spectra from different Instruments. For comparison, a virtual absorption spectrum was formed from the previously extracted

features m_i. The relevant feature for the protein concentration also determined in the competition literature is the small peak at $1016\ nm$. The information about the protein content is thus in a small hidden peak [IAB+17].

The determined features m_i of all 6 spectrometers for all 248 samples are shown in a three-dimensional feature map in a section around the relevant protein peak (Fig. 4.3). The wavelet coefficient is normalised by referencing to a protein independent peak. The presented method found the relevant peak in all 1488 spectra. The figure also clearly shows, that the width s_i and wavelet coefficient $\Gamma_\psi(\lambda_i, s_i)$ of the protein peak increases with the protein content.

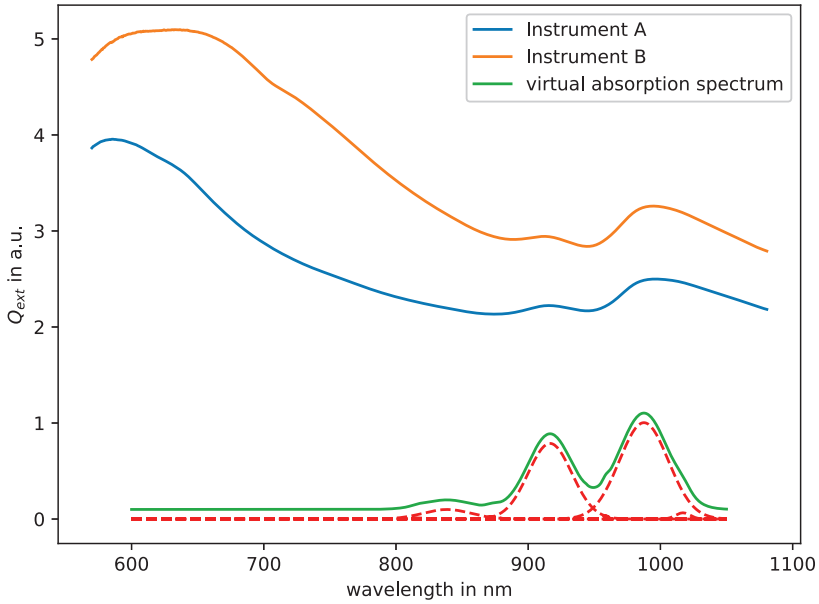


Figure 4.2: The spectra of an identical sample recorded by two spectrometers models from different manufacturers A and B. In the lower part is shown as an example a virtual absorption spectrum, which was determined from the previously extracted features m_i

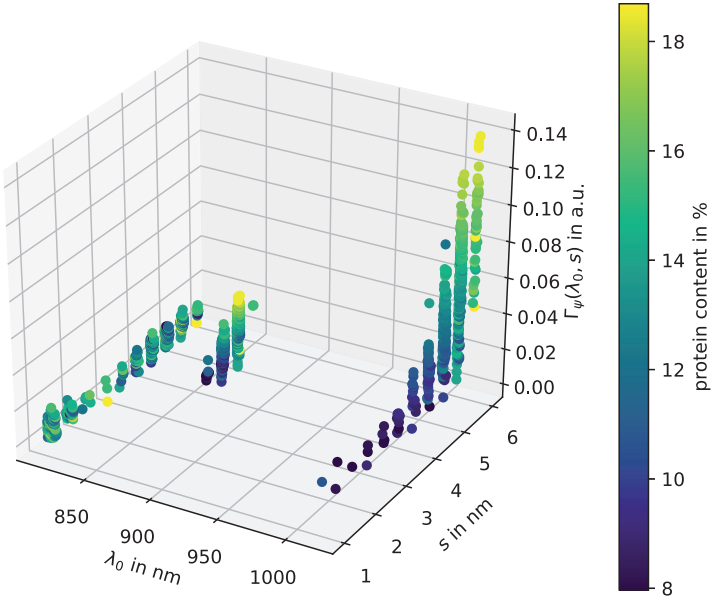


Figure 4.3: The three-dimensional feature map shows the parameters of the features extracted from the training set. The protein content of the sample is shown in color. For better visibility of the protein content, the presentation is limited to three small features.

5 Summary

The presented method is suitable for the feature extraction of superimposed absorption signals. The position and width of a peak are independent of the scaling of the signal strength and are therefore suitable for a robust material identification, regardless of the sample geometry and the measuring device used. The information of the signal intensity is included in the wavelet coefficient and offers the possibility to quantify an ingredient. The representation of the spectral features m_i as a list of triplets $(\lambda_i, s_i, \Gamma_\psi(\lambda_i, s_i)$ can be created for measurements of sensors of different types. Classification and regression models based on evaluation of the triplets are thus invariant with respect to the sensor used.

The knowledge of the position and width of the absorption features also allows further evaluations. Optical filters can be selected based on the individual features. In addition, components of the elastic scattering parameters can be determined from the residuum of the spectral signature after deduction of the absorption properties. In hyperspectral imaging, feature extraction can be used for compression.

Bibliography

[CDL02] A. J. Cox, Alan J. DeWeerd, and Jennifer Linden. An experiment to measure Mie and Rayleigh total scattering cross sections. *Am. J. Phys.*, 70(6):620–625, 2002.

[Dem10] Wolfgang Demtröder. *Expirmentalphysik 3, Atome, Moleküle und Festkörper*. Springer-Lehrbuch. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[DWKR16] Anshuman J. Das, Akshat Wahi, Ishan Kothari, and Ramesh Raskar. Ultra-portable, wireless smartphone spectrometer for rapid, non-destructive testing of fruit ripeness. *Sci. Rep.*, 6(September):1–8, 2016.

[FWT+02] Robert N. Feudale, Nathaniel A. Woody, Huwei Tan, Anthony J. Myles, Steven D. Brown, Joan Ferre, and Joan Ferré. Transfer of multivariate calibration models: a review. *Chemom. Intell. Lab. Syst.*, 64(2):181–12, 2002.

[IAB+17] Benoit Igne, Md Anik Alam, Dongsheng Bu, Pierre Dardenne, Hanzhou Feng, Ali Gahkani, David W Hopkins, Shikhar Mohan, Charles R Hurburgh, and Cathleen Brenner. Summary of the 2016 IDRC software shoot-out. *NIR news*, 28(4):16–22, 2017.

[LGGFR17] Mercedes G. López, Ana Sarahí García-González, and Elena Franco-Robles. Carbohydrate analysis by NIRS-chemometrics. In *Dev. Near-Infrared Spectrosc.* InTech, mar 2017.

[Mal89] Stephane G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

[Mie08] Gustav Mie. Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Ann. Phys.*, 330(3):377–445, jan 1908.

[MNHG96] C. R. Mittermayr, S. G. Nikolov, H. Hutter, and M. Grasserbauer. Wavelet denoising of Gaussian peaks: A comparative study. *Chemom. Intell. Lab. Syst.*, 34(2):187–202, 1996.

[Mor83] J. Morlet. Sampling theory and wave propagation. In C. H. Chen, editor, *Issues in Acoustic Signal — Image Processing and Recognition*, pages 233–261, Berlin, Heidelberg, 1983. Springer Berlin Heidelberg.

[NW84] KH Norris and PC Williams. Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat. I. Influence of particle size., 1984.

[RDC17] Giovanni Rateni, Paolo Dario, and Filippo Cavallo. Smartphone-based food diagnostic technologies: A review. *Sensors (Switzerland)*, 17(6), 2017.

[RvdBE09] Asmund Rinnan, Frans van den Berg, and Soren Balling Engelsen. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, 28(10):1201 – 1222, 2009.

[SG64] Abraham Savitzky and Marcel J.E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, 36(8):1627–1639, 1964.

Adaptive Measurement Method for Area Chromatic Confocal Microscopy

Ding Luo

Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT), Germany
ding.luo@kit.edu

Technical Report IES-2017-03

Abstract: Although well known for its depth discerning capability, conventional confocal microscopy has limited application due to its slow scanning speed. From Nipkow disk to various programmable spatial light modulator, various research has been conducted with the aim of improving the speed of confocal microscopy. Nevertheless, the fundamental conflict between axial sensitivity and lateral density remains unsolved. In this report, a novel adaptive measurement method is proposed based on iterative refinement of the axial measurement as well as condensation of the lateral measurement grid. Initial experimental investigation has shown overall good measurement result with a specific type of artifacts due to inaccurate estimation in earlier measurement stages. Despite of this problem, the proposed system and the accompanying algorithms have shown great potential in improving the measurement speed of area chromatic confocal microscopy.

1 Introduction

Conventional scanning confocal microscopy suffers from a relatively slow measurement speed due to its requirement for mechanical scanning, which largely limits its application in various fields. To tackle this problem, Egger et al. first proposed to utilize the Nipkow disk to generate a moving array of measurement

locations in order to accelerate the scanning process [EP67]. Recently, more advanced disk pattern has been designed to be coupled with structured illumination technology, in order to achieve superresolution imaging [HO15].

With the development of new optical components and computer technology, this idea of using an array of measurement locations has transformed into an important field of research, i.e. programmable array microscopy (PAM). Programmable array microscope refers to a family of imaging systems where a spatial light modulator (SLM) is applied to dynamically change the patterns of illumination and/or detection. With the target of eliminating lateral mechanical scanning, different SLMs have been applied, including digital mirror device (DMD) [HVG+99, CRS15], liquid crystal on silicon (LCoS) [HCT+07, KDP+14], and polymer-dispersed liquid crystal (PDLC) [CS17].

Despite the improvement achieved through these developments, one fundamental problem remains unsolved. The unique depth discerning capability of the confocal technology originates from the fact that the light which is not focused on the object is distributed to the adjacent area, thus dramatically reducing the reflected light collectable to the confocal pinhole. Such a principle intrinsically demands larger numerical aperture (NA) to generate highly focused spot, in order to achieve better axial sensitivity and lateral resolution, which has not been a problem for conventional scanning confocal microscopy. Nevertheless, for array scanning microscopy, realized through whether mechanically scanned disk or spatial light modulator, the blurred illumination spot of one measurement location quickly generates crosstalk over its adjacent measurement locations. This leads to an inverse relationship between the minimum allowable pitch of the measurement array and the NA of the system as well as the axial measurement range.

This report aims to provide a potential solution for this problem through an adaptive measurement strategy based on a particularly dynamic hardware setup, which will be discussed in details in the following sections.

2 System Setup and Calibration

The proposed system is composed of two components, i.e. a programmable light source and a DMD-based programmable array chromatic confocal microscope. Due to its nature of adaptability, the system is denoted as AdaScope.

2.1 Programmable Light Source

The programmable light source is based on two-dimensional dispersion of a white light laser. As illustrated in Fig. 2.1, the laser is firstly dispersed horizontally by a prism and then dispersed vertically by the echelle grating to achieve very high overall dispersion. A digital mirror device is used to select the desired wavelengths, which is collected by the output liquid light guide. The system is capable of generating light spectrum in the range of $480\,\mathrm{nm}$ to $680\,\mathrm{nm}$. More details regarding the programmable light source can be found in the corresponding paper [LTLB17].

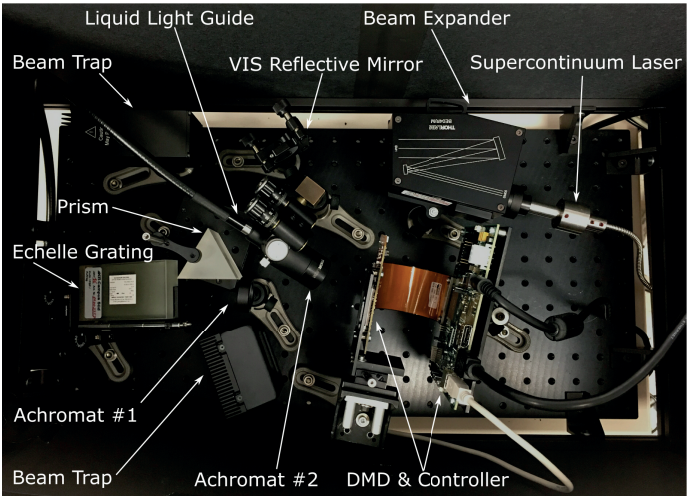


Figure 2.1: Setup of programmable light source.

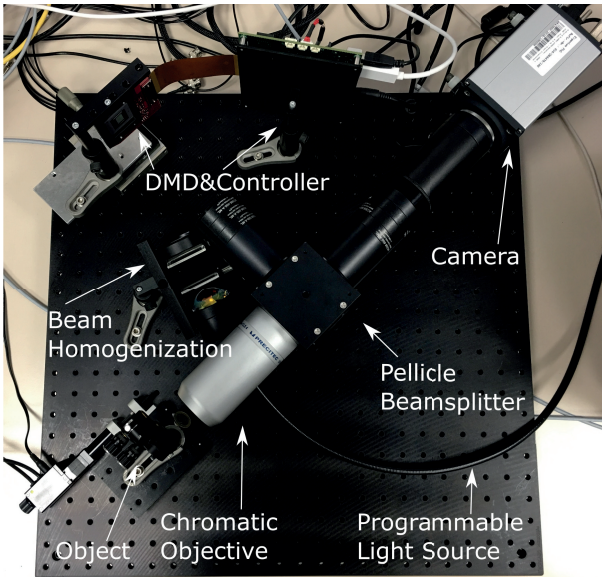


Figure 2.2: Setup of programmable array microscope.

2.2 DMD-based Programmable Array Microscope

In the microscope setup, light coming from the programmable light source is first homogenized and projected onto the DMD, which acts as an array of secondary sources. Illumination light is projected onto the object using an objective (Precitec CLS4) with designed chromatic separation along the optical axis. The reflected light travels through the same objective and is collected by the camera. The sCMOS camera (Andor Zyla 5.5) in the system has very good signal to noise ratio and color depth (dynamic range) but not a very fast speed. As will be later discussed, since the speed of the programmable light and the DMD in the microscope are both very fast, the frame rate of the camera will be a major limiting factor that has to be considered when designing the measurement algorithm. The measurement area is $5.4\,\mathrm{mm}$ by $3.0\,\mathrm{mm}$ laterally and the depth measurement range is $4.6\,\mathrm{mm}$ for a wavelength range of $200\,\mathrm{nm}$.

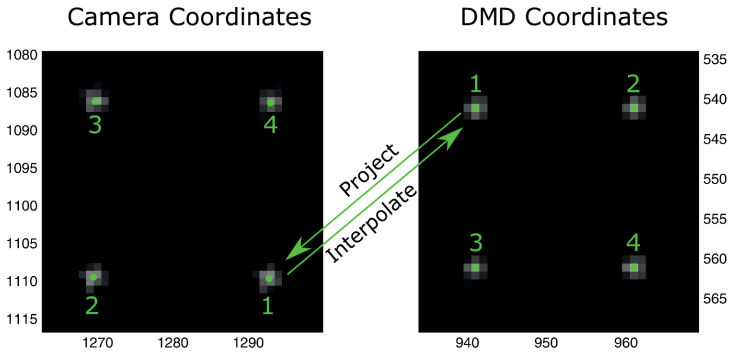


Figure 2.3: Camera calibration for AdaScope.

2.3 Camera Calibration

As a microscope system, the illumination homogenization and projection system, the DMD and the camera arm all have to be aligned very accurately for the confocal principle to work. After the alignment, camera calibration is implemented. Since the system is designed to be telecentric, the camera is only calibrated for one wavelength ($555\,\mathrm{nm}$) where a registration is made between the camera coordinates and the DMD coordinates. The registration toward the object / world coordinate system is not considered at the moment. All measurements with AdaScope shown in this report are in the DMD coordinate system by first projecting the DMD coordinate to the camera coordinate system and then making an interpolation, as shown in Fig. 2.3.

As demonstrated in Fig. 2.4, due to the large NA of the microscopic objective, when a flat mirror serves as the target object, the blurred spot due to defocus covers a large area even at a small distance. The image demonstrates the burred spot when the mirror is located at a distance of $95.25\,\mu\mathrm{m}$ from the focal plane. This corresponds roughly to a wavelength shift of $4\,\mathrm{nm}$ in the chromatic confocal scan. It can be seen that the crosstalk reaches more than a distance of 10 pixels already. To fully suppress the effect of crosstalk, a minimum pitch of 20 pixels is required. To scan through 200 wavelength steps, 80000 frames have to be taken, which costs roughly $0.75\,\mathrm{h}$ of acquisition time. If certain levels of crosstalk can be tolerated, a pitch of 10 pixels can be taken, which leads to an acquisition

time of 11 minutes. Even in this case, the speed of measurement still cannot be considered to be practical for real industrial applications.

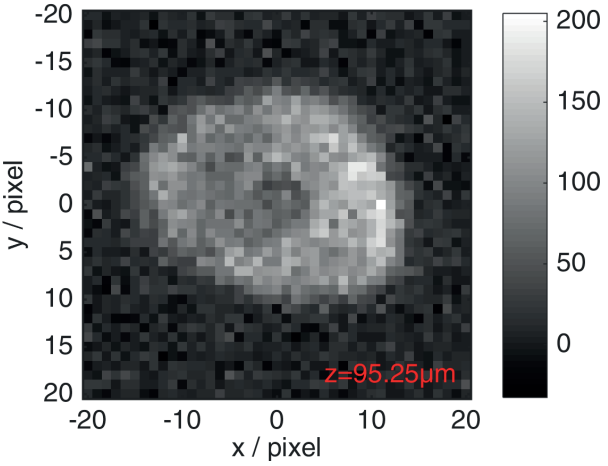


Figure 2.4: Defocus blurring.

3 Adaptive Measurement Method

To further accelerate the speed of area chromatic confocal measurement, an adaptive grid resizing algorithm has been developed. The idea originates from the observation that the uncertainty of the chromatic confocal measurement is in fact coupled with the lateral density of measurement locations, as shown in Fig. 3.1. When little information of the measurement locations is gathered, the crosstalk could potentially be very large and therefore a larger distance between adjacent points is required. As the measurements at each point become more and more accurate, the possibly generated crosstalk also gets smaller which allows for a denser measurement array.

Based on this observation, the measurement is conducted in several iterations. In each iteration, measurements with limited accuracy are made for all positions through array scanning with a fixed pitch distance. Based on the result from one iteration, more refined measurements are made with a denser grid in the next iteration.

3.1 Axial Measurement Refinement

In this iteration, a two-channel linear measurement system is applied to each measurement location. The two measurement functions are two ramp-shaped functions in opposite directions. To measure the axial location of the corresponding chromatic confocal peak, illuminations with spectra in the shape of the measurement functions are applied and the corresponding images are captured. As 1st order Bernstein polynomials, these functions have the nice property that the corresponding linear transformation maintains the centroid of the original signal. Therefore, the centroid of the chromatic confocal peak can be estimated with very fast computation:

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} \mathbf{f_1}^T \\ \mathbf{f_2}^T \end{bmatrix} \mathbf{g}$$

$$centroid(\mathbf{g}) = centroid(\mathbf{m}) = \frac{m_2}{m_1 + m_2}$$

where $\mathbf{g}$ represents the original confocal signal, $\mathbf{m}$ denotes the measurement, $\mathbf{f_1}$ and $\mathbf{f_1}$ represent the illumination spectra.

There are several reasons for using such a linear measurement system. Firstly, since more than one iterations are performed, each iteration must be very efficient in terms of the number of frames taken. Secondly, the crosstalk at a fixed distance should be proportional to the measurement range. This means that as the location of the object becomes more certain, the crosstalk should become smaller. Lastly, the uncertainty should be inversely proportional to the measurement range. This means that for a smaller measurement range, the sensitivity should be higher.

All these properties are achieved by iteratively reducing the wavelength range of the illumination according to the previous estimation, such as illustrated by

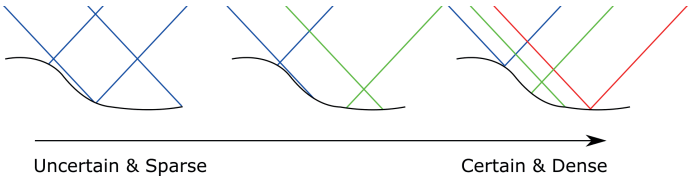


Figure 3.1: Coupling of axial measurement uncertainty and lateral measurement density.

Fig. 3.2. Suppose the position of the object is represented by the arrow. In the first iteration, the camera takes two frames with the two illumination spectra covering the complete wavelength range. Based on estimation result from the first iteration, which is not extremely accurate, the object is determined to be in the top half of the measurement range. In the second iteration, the AdaScope makes measurement in the new measurement range with two similar illumination spectra. This appears to be like a binary search, but if the measurement in each iteration is accurate enough, the search process can be much faster. For example, a direct jump from iteration #1 to iteration #3 will also be possible.

Apparently this method is not very sensitive and is not robust against the noise due to the limited number of linear measurement channels, but it should be sufficient to bound the measurement range to a certain level for the next iteration.

3.2 Lateral Grid Condensation

As mentioned previously, in each iteration, the measurement density is also increased accordingly. As shown by the example in Fig, 3.3, in iteration #1 with a pitch of 20 pixels, grid has to be scanned 20 by 20 times, and in each time, the system makes two measurements using the corresponding illumination spectra. In the next iteration, the density of the grid can be increased depending on how much the new measurement range is bounded.

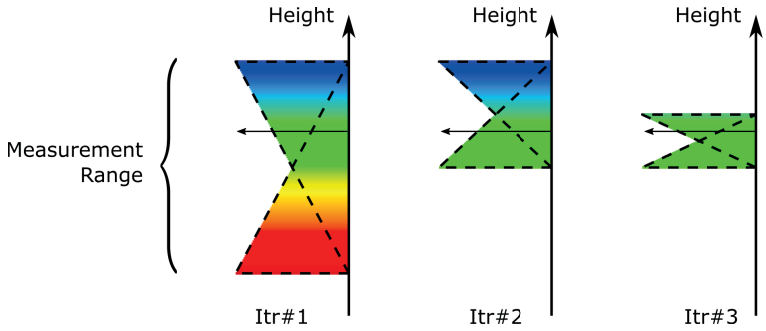


Figure 3.2: Iterative refinement of axial measurement.

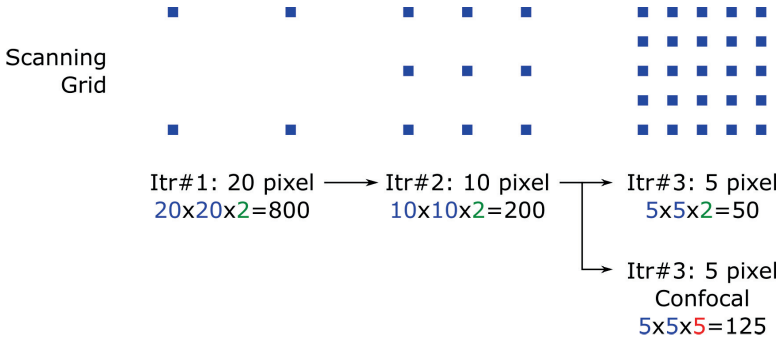


Figure 3.3: Iterative condensation of lateral grid.

At a certain iteration, based on the estimation uncertainty from previous iteration, the measurement process can be switched to a localized chromatic confocal measurement centered around the previous estimation result, in order to get more accurate measurement result.

3.3 Hardware Triggering

As an example, the triggering diagram for the second iteration as well as the corresponding illumination spectra are illustrated in Fig. 3.4. Since the camera is the slowest component, it serves as the master which triggers the spectral DMD in the programmable light source. This DMD displays several patterns corresponding to several illumination spectra. Each spectral DMD patterns triggers its corresponding spatial DMD pattern in the microscope. Based on estimation from the first iteration, all points are already bounded to either the top half or the bottom half of the complete measurement range. For each measurement grid, two frames are captured. Within each frame, two spectra are projected to two different spatial patterns. In the second frame, the spatial patterns are repeated but the spectra are different. This process is then repeated pitch^2 times for complete measurement of this iteration.

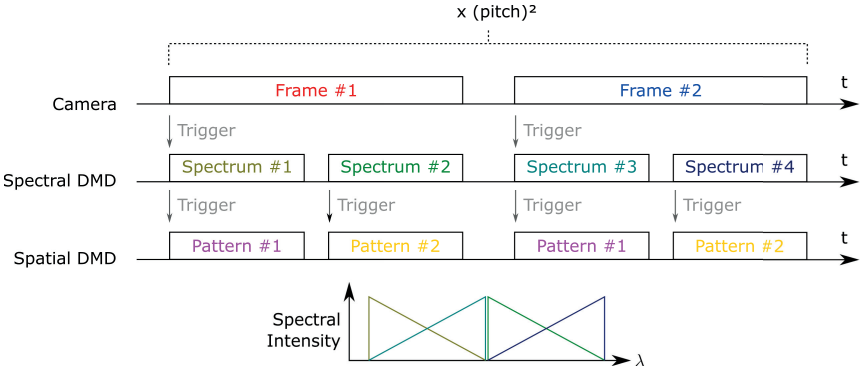


Figure 3.4: Exemplary triggering diagram of iteration #2.

4 Experiment and Analysis

To investigate the feasibility of this adaptive measurement method, a test measurement is conducted with a one euro coin serving as the target. For proof of concept, the wavelength range is limited to $530\,\mathrm{nm}$ to $580\,\mathrm{nm}$, which corresponds to a measurement range of $1.15\,\mathrm{mm}$.

For the adaptive scan, in each iteration, two images are captured with different illumination spectra and spatial patterns. Fig. 4.1 shows the raw measurement data from the first iteration with their corresponding illumination spectra. It can be seen that the image with first illumination spectrum is much brighter than the second one. This indicates that the centroid of the chromatic confocal peak is closer to the focus position of $530\,\mathrm{nm}$.

After the first iteration, a rough estimation of the height can be performed based on the two-channel linear measurement principle, which is then binarized into two levels. As can be seen in Fig. 4.2, for most of the area, the height lies in the upper half of the current measurement range.

After the second iteration with a denser grid, more accurate estimation is performed, which generate four measurement range divisions (Fig. 4.3).

After iteration #2, the system directly makes localized chromatic confocal measurement with five wavelength steps with a step size of $1\,\mathrm{nm}$. These five steps

are centered around the estimation result from iteration #2. In this iteration, the pitch is further reduced to 5 pixels. As shown in Fig. 4.4, estimation of the height and the intensity (texture) is performed through Gaussian fitting on the localized measurements.

Although the overall estimation is correct, there are apparent artifacts all across the measurement area. The major reason for this kind of artifact is the inaccurate estimation result before the localized chromatic scanning. Since the iteration of the localized scan is initialized based on previous estimation, when the starting point is already too far away from the actual chromatic peak, two adverse effects could happen. On one hand, the five wavelength scanning steps are not enough to cover the peak position of the chromatic confocal signal. On the other hand, when the scanning wavelength steps are too far from the peak, the generated crosstalk for the adjacent locations will no longer be tolerable for the selected pitch distance of the measurement grid. To avoid such artifacts, more accurate estimation from the linear measurement stage must be achieved in order to correctly initialize the localized chromatic scanning, which will be the key part in future research.

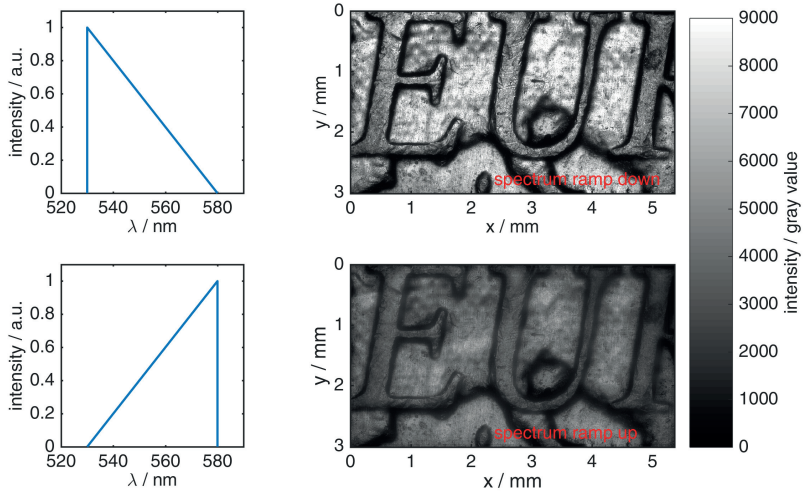


Figure 4.1: Illumination spectra (left) and raw measurements (right) from iteration #1.

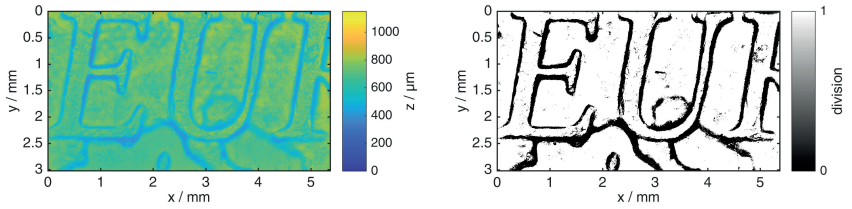


Figure 4.2: Estimation (left) and binarized levels (right) from iteration #1.

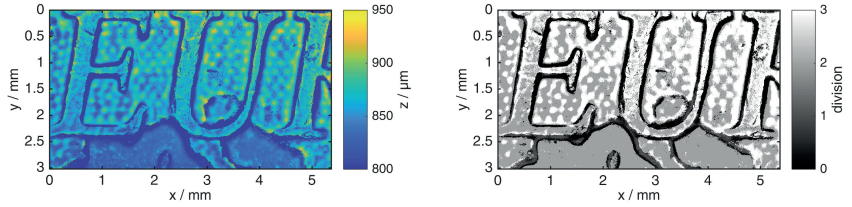


Figure 4.3: Estimation (left) and binarized levels (right) from iteration #2.

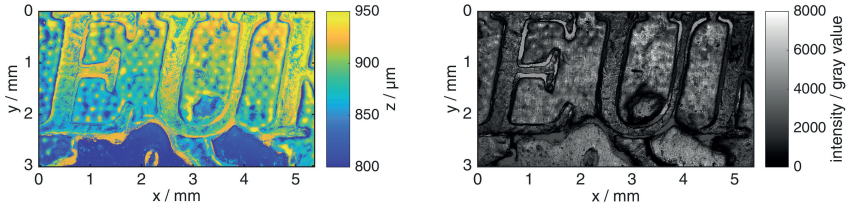


Figure 4.4: Height estimation (left) and intensity (right) from iteration #3.

5 Conclusion

This report presents a novel approach to the problem of area chromatic confocal microscopy. The hardware setup consists of two fundamental components. One part is a programmable light source based on a white light laser. The other part is a DMD-based programmable array chromatic confocal microscope. The combination of both, namely the AdaScope, provides an adaptive platform which

can be utilized for 3D measurement of reflective/defusing samples. An adaptive measurement method is developed based on this system setup, which iteratively improves the axial measurement uncertainty and lateral measurement density. Although initial experimental investigation has revealed certain artifacts in the measurement result mainly due to inaccuracy in the linear measurement iterations, it is believed that such approach has great potential in increasing the measurement speed compared to a naive fixed grid scanning. This could lead to applications of the area chromatic confocal measurement technology in real industrial settings.

Acknowledgement Research conducted in this report is financed by the Baden-Württemberg Stiftung gGmbH.

Bibliography

[CRS15] Nadya Chakrova, Bernd Rieger, and Sjoerd Stallinga. Development of a DMD-based fluorescence microscope. *Proc. SPIE*, 9330:9330–9330–11, 2015.

[CS17] Ting-Jui Chang and Guo-Dung J. Su. A confocal microscope with programmable aperture arrays by polymer-dispersed liquid crystal. *Proc. SPIE*, 10376:10376–10376–7, 2017.

[EP67] M. David Egger and Mojmir Petran. New reflected-light microscope for viewing unstained brain and ganglion cells. *Science*, 157(3786):305–307, 1967.

[HCT+07] Guy M. Hagen, Wouter Caarls, Martin Thomas, Andrew Hill, Keith A. Lidke, Bernd Rieger, Cornelia Fritsch, Bert van Geest, Thomas M. Jovin, and Donna J. Arndt-Jovin. Biological applications of an LCoS-based programmable array microscope (PAM). *Proc. SPIE*, 6441:6441–6441–12, 2007.

[HO15] Shinichi Hayashi and Yasushi Okada. Ultrafast superresolution fluorescence imaging with spinning disk confocal microscope optics. *Molecular Biology of the Cell*, 26(9):1743–1751, 2015.

[HVG+99] Q. S. Hanley, P. J. Verveer, M. J. Gemkow, D. Arndt-Jovin, and T. M. Jovin. An optical sectioning programmable array microscope implemented with a digital micromirror device. *Journal of Microscopy*, 196(3):317–331, 1999.

[KDP+14] Sharon V. King, Ana Doblas, Nurmohammed Patwary, Genaro Saavedra, Manuel Martínez-Corral, and Chrysanthe Preza. Implementation of PSF engineering in high-resolution 3D microscopy imaging with a LCoS (reflective) SLM. *Proc. SPIE*, 8949:8949–8949–7, 2014.

[LTLB17] Ding Luo, Miro Taphanel, Thomas Längle, and Jürgen Beyerer. Programmable light source based on an echellogram of a supercontinuum laser. *Applied Optics*, 56(8):2359–2367, 2017.

Deterministic Industrial Network Communication: Fundamentals

Ankush Meshram

Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT), Germany
ankush.meshram@kit.edu

Technical Report IES-2017-04

Abstract: Industrial networks came into existence with the third industrial revolution to support manufacturing and automation. Over the years, there has been technical advancement in different aspects of networking technologies in order to make production and governing automation efficient and intelligent. This also brought along advancing threats leading to the need of advancements in counterattacking or prevention methods. However, to contribute in challenging the Advanced Persistent Threats (APTs) the understanding of the fundamentals of industrial communication is needed. Determinism is at the core of automation, hence this report comprehends various literature sources on the industrial network communication strategies to achieve deterministic industrial network communication.

1 Introduction

Networks have become an integral part of manufacturing over the years replacing point-to-point communications at all levels. At lower levels in factory infrastructure, networks provide higher reliability, visibility and diagnosability and enable capabilities such as distributed control, diagnostics, safety and device interoperability. At higher levels, networks can leverage Internet services to enable factory-wide automated scheduling, control, and improve data storage and visibility. Industrial networks were introduced considering varying requirements of factory automation, distributed process control, home automation, and of critical

Infrastructures such as energy distribution as well as transportation. Appropriate networking technology evolved simultaneously within the application field. The three major influences identified for industrial network evolution [Zur14] are:

- Communication engineering for data transmission over large telephone networks in telecommunication sector.
- Instrumentation and measurements systems with parallel buses to account for limited data processing speed and real-time requirements for synchronization.
- Computer science with high-level communication protocol designs, such as WANs and LANs, leading to gradual change of analog to digital systems in telecommunication sector.

Fieldbus systems were the landmark in industrial networks evolution for automation, which replaced traditional expensive point-to-point cabling of devices to central control room. It brought concepts of decentralization, modularity to extend installations, and communication between intelligent devices, capable of data preprocessing, for transferring process data, and parameterization and configuration purposes. The idea of computer-integrated manufacturing (CIM) comprehended the structure of information flow required for automation in a hierarchical model — to create a transparent, multilevel network — called automation pyramid [SSKD11]. It comprised of 5 or more levels, in order of lowest level at the bottom to highest level at top: Field level (sensor/actuator), Process level (Programmable Logical Controller (PLC), Human Machine Interface (HMI), Cell level (Operator station), Factory level (Manufacturing execution systems (MES)), Company level (Enterprise resource planning (ERP)). Fieldbuses populated the field, process and cell levels while bridging the gap between lower levels which traditionally consisted of point-to-point connections to higher level networks. The distinction between lower level and higher level networks of the automation pyramid is maintained by fieldbus systems. The popularity of Ethernet as the LAN technology in automation and its penetration of all levels of this pyramid to process level is likely to replace mid-level fieldbus systems. Industrial Ethernet is resulted in reduction of the levels in the automation hierarchy, and ultimately to flattening out of pyramid to at most three or two levels.

In further sections, first we will build on communication fundamentals to understand industrial communication paradigms. This is followed by section 3 on

relation between industrial communication and determinism. In section 4, we elaborate on the foundations of deterministic industrial protocols and ending the report with short summary.

2 Fundamentals of Industrial Communication

2.1 Communication Layers

Introduction of the ISO/open system interconnection (OSI) seven-layers reference model for data communication has been the foundation for development of complex communication protocols [Zim80, DZ83] (Figure 2.1). There are three important concepts to understand before dwelling into layers:

1. *Protocol*, is a set of rules and convention that communication layer N of open system must confer to communicate with layer N of another open system. Rule sets of each layer define the respective layer protocol.
2. *Service*, defines the functionality of services offered by one layer (service provider) to layer above it (service user). The OSI model doesn't enforce the way services are implemented.
3. *Interface*, specifies interface between layers with services offered by the lower layer to the upper layer. It also defines access methods with parameters and what results to expect.

An application system sends information to another system through packaging data at top layer and requesting services of layer below to transmit data, repeating until lowest layer. On the way down the layers, the data of the application process are augmented by layer-specific data needed to execute the respective protocols. These data are typically address and control information that is mostly combined in a protocol header. In addition, the data may be segmented into individual packets to match the allowed maximum packet size for a given layer. This way, the number of bits being actually transmitted can be significantly larger than the pure user data provided by the application process,and the communication overhead can be substantial. The receiving system strips additional information of the peer layer to recover data for the application process. The layers of the OSI model, bottom-up, are briefly described next.

- Layer 1, or Physical Layer, presents all mechanical, physical, optical, electrical, and logical properties of the communication system to upper layers necessary for transferring data frames.

- Layer 2, or Data Link Layer, is responsible for data frame formation from bits with frame's coding and checking for transmission errors (via Cyclic Redundancy Check). It is subdivided into the logical link control (LLC) and medium access control (MAC). LLC takes care of error detection mechanism and sets up the connection to layer 3. MAC links to layer 1 and controls who is able to transmit when.

- Layer 3, or Network Layer, establishes routing paths between origin and destination nodes of end-to-end connections while assigning special target addresses. The transmission paths are optimized to reduce congestion in the presence of multiple physical transmission mediums with varying transmission speeds and not exceed maximum allowed delay.

- Layer 4, or Transport Layer, sets up the end-to-end connection and splits up the data in small numbered packets whenever either the data size is too big or transmission times are long. The peer layer on the receiving system takes care of recombining the individual packets in the right order.

- Layer 5, or Session Layer, synchronizes the communication between participating systems while handling the authentication and identification of devices. To perform its synchronization task effectively it introduces any synchronization markers to resume after communication breakdown.

- Layer 6, or Presentation Layer, codes the transmitting data and its interpretation on the receiving system. It interprets the syntactic bit sequence of data into a character and its semantic meaning, such as currency or physical units.

- Layer 7, or Application Layer, provides the interface between the application and the communication unit for transparent representation of communication. It defines the procedures or protocol processes of various application functions for calling up data, file transfer, etc. It is designed in such a way that a system accesses information through its communication unit without the need to know the functions of underlying layers.

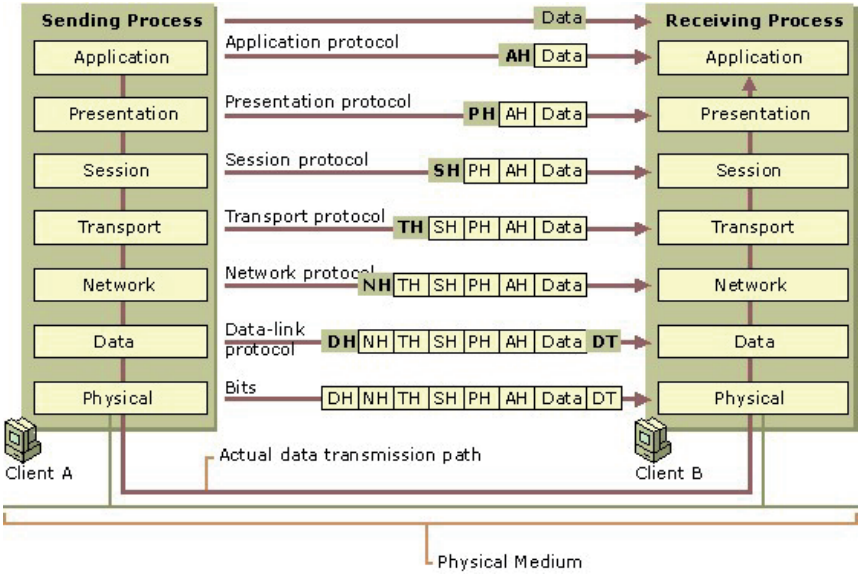


Figure 2.1: ISO/OSI stack with frame headers [Shi12].

Following the hierarchical OSI model, it is possible to set up complex communication system between heterogeneous systems on different layers. Through the use of repeaters, one can overcome the limitations of a given physical layer. The interconnecting devices share a common data link layer. Bridges interconnect different networks by translating data and protocols on layer 3. Routers link networks on layer 4, whereas gateways (especially, application layer gateways) interconnect entirely different communication systems on the application layer.

2.2 Communication Types and Services

Two distinct techniques are used in data communications to transfer data – connection-oriented and connection-less [Zim80]. A connection-oriented method (virtual circuit service) requires a session connection be established before any data can be sent. This method guarantees that data will arrive in the same

order. Connection-less method (datagram service) doesn't require a session connection between sender and receiver. There is no guaranteed data arrival however it is useful for periodic burst transfers. TCP (Transmission Control Protocol) is a connection-oriented transport protocol, while UDP (User Datagram Protocol) is a connectionless network protocol, both operating over IP at the Transport Layer.

The interface between neighboring service provider and service user layers is called service access point (SAP). The user data of layer N to be sent across the network for its peer layer is encoded as service data unit (SDU) and passed on to lower layers in the hierarchy for further processing. The communication between two peer layers is governed by the rule sets (protocol) that can only be understood by them. This information is added to SDU with interface control information (ICI) and protocol control information (PCI) to form a cohesive protocol data unit (PDU).

There are four primitive operations through which interaction between layers occur – Request, req, Indication, ind; Confirmation, con; and Response, res. Service user layer invokes request and response while resulting confirmation and indication originate from corresponding service provider. The combination of these primitive operations can be categorized into 3 major service categories – unconfirmed service, confirmed service, and acknowledged service. The unconfirmed and confirmed service comprise of a request, an indication and a confirmation, whereas acknowledged service comprise all the operations.

2.3 Communication Mechanisms

The multitude of application domains of automation systems have different timing and consistency requirements which varies too within the application areas such as manufacturing, process automation, etc [Tho05].

The timing behavior of a technical process can be conceptualized as either state-based or event-based. In the state-based approach, the status of internal state variables (temperature, pressure) of the process are continuously sampled and transmitted in discrete-time for continuous process control and monitoring. The event-based approach transmits data only when the process state changes and are well-suited for discrete processes or subprocesses which can be modeled as a state machine.

The traffic in industrial networks could be periodic or aperiodic based on how often process data is accessed. Periodic, or cyclic, traffic follows time-slot-communication strategy where each state variable is assigned a dedicated slot in the available bandwidth based on the a priori information of its data generation rate or sampling time. The update rate in a periodic traffic is usually dynamic adapting to the sampling rate demand of current process state information and the data exchange is connection-less. The periodic data is handled through buffers following the FIFO (first-in-first-out) structure where older values are overwritten by latest data. Aperiodic, or acyclic, traffic is generated on demand in an event-based manner and transmitted when free communication bandwidth is available or idle time is reserved between time slots of periodic traffic. Aperiodic data is connection-oriented where acknowledgment is used to allow for re-transmission of lost configuration data. Queues handle aperiodic data where messages are not overwritten and no new data is accepted when the queue is full.

Based on the consistency of accessing information in automation systems, the traffic data can be classified as continuously updated process data and on-demand parameterization data. The process data are real-time data and could be periodic or aperiodic requiring strict delivery timing to be meaningful for process control. The process data in the events of transmission errors could be reconstructed from historical data via interpolation. However, aperiodic process data additionally requires that no data lass occurs or atleast detected in due time through appropriate mechanisms. The parameterization, or configuration, data are non-real-time and usually aperiodic though session information, authentication information or updated communication parameters are transmitted periodically or quasi-periodically. The configuration data contains necessary information to set up or adjust the operation of automation system and needs guaranteed consistent delivery across the system.

2.4 Network Topologies

The star topology was the default wiring structure in automation before fieldbus was introduced. The PLC is at the center connected to I/O elements individually. The line, or bus, topology evolved as most efficient replacement to star-like point-to-point cabling and was quickly adapted network topology. The nodes are all connected in one single line. In the ring topology, nodes are arranged one after another in the form of a chain where each nodes has two independent interfaces

for input and output. It is very fast and deterministic method to exchange data with low jitter (variance of time delay) as nodes don't need explicit addressing. A variant of the ring topology is daisy-chain structure where nodes are cascaded like a string of pearls. In the tree topology, nodes are arranged in hierarchical composite network structure where each node could be a root for a lower-level segment. The root nodes usually have routing capabilities, so that the data traffic can at least partly be confined to individual areas of the network. In the mesh networks, there exists multiple paths between nodes through the network. It requires appropriate routing strategies to keep messages from circling in the network and causing congestion.

2.5 Medium Access Control

The topology being used by the networking technology influences the selection of the medium access control (MAC) method, or vice versa [Zur14].

The data transfer mechanism can be classified into two – single-master (or master-slave) and multimaster. In the single-master approach follows centralized communication architecture where the master either retrieves data from its slaves following request-response communication or synchronizes time slots with slaves to send their data. Such networks are usually single-segment structures with limited size and found at lowest levels of automation pyramid. Within the multimaster approach, participating nodes have equal rights over the communication medium and share it in a democratic way. These networks are found on the middle level of automation pyramid.

Time division multiple access (TDMA) is the actual MAC strategy used over multiplexing methods such as frequency division multiple access (FDMA), code division multiple access (CDMA), or space division multiple access (SDMA) for industrial communication. In TDMA, the network nodes shares the bandwidth and communicate sequentially. The basic methods of multiple access follows either centralized approach by polling or time-slot-based techniques, or in decentralized way by token passing or random access methods.

Polling is a master-slave mechanism where a slave node sends information only when explicitly called upon by the master node. In the network, alternate poll messages from master to each of its slave and responses from them is observed. Polling is strictly cyclic where the master polls all the slaves sequentially and

restarts the cycle. It's cyclic behavior suits well for periodic traffic where process variables are polled equidistantly. Polling polls data data either by explicit node addressing or process variable identifier irrespective of device generating the values. The later variant is also called as central polling. Strict polling doesn't offer functionality for aperiodic traffic and slaves cannot become active themselves to send event of an alarm condition. However, there exist alternate mechanisms to rectify these disadvantages.

Time-slot-based method divides available transmission time on the medium into distinct slots where slaves access the medium at its assigned time slot. The cyclic polling in its essence too partitions the polling cycle into time windows, however in time-slot-based method slaves can send the data themselves without a request from central master. Time-slot-based methods are mostly referred as TDMA. Synchronous TDMA equally distributes time slots whereas asynchronous TDMA dynamically distributes time slot according to amount of data to be sent. Aperiodic traffic is accommodated between the cyclic time slots. Based on the mechanism incorporated to synchronize slots there are two variants of TDMA – centralized and decentralized. In the centralized approach, a dedicated master sends some sort of synchronization message at the start of the cycle followed by nodes exchanging data in their pre-assigned time slots. On the contrary, in the decentralized approach all the nodes synchronizes themselves without explicit node to initiate cycle. Either explicit clock synchronization mechanisms or set of timers that set operation to a stable state state are used by the nodes.

Token Passing (TP) method of medium access is based on a special piece of information, called token, passed on between peer network nodes and only the node possessing it can initiate the data transfer. A set of rules ensure its fairness and its detection when lost or duplicated. Compared to time-slot mechanisms, when a node possessing the token doesn't have any data to send it passes the token on to the next node thus saving time. TP can be implemented either explicitly through a dedicated short message or implicitly through distributed, synchronized access counters (ACs) included in all nodes. The explicit form of TP uses target token rotation time (T_{tr}) to enforce the maximum time duration to posses the token. The implicit TP uses two counters, ACs and Idle Bus Bit Period Counter (IC), included in every master to simulate the token.

Another peer-to-peer communication based medium access method is random access where a network node tries to access the communication medium whenever it wants to without any imposition. This is also called carrier sense multiple access (CSMA). However, the major drawback of this approach is collisions when several nodes try to send data at same time, even if they noticed idle communication line before sending. The variants of CSMA deals with collisions in different ways to avoid bandwidth wastage and communication delays. In CSMA-CD (collision detection), collisions are detected by sending nodes which aborts the data transfer and wait for a random time before trying to send again. In *p*-persistent CSMA variant , the waiting time depends on the value of probability (*p*) that the node will try again in a certain time interval after collision. The probability of each node is adaptable to the estimation on its backlog and the monitored network load. The widely used CSMA-CA (collision avoidance) variant uses asymmetric symbols for coding the bits on the communication line, so that when two different bits are sent at a time, the dominant one wins over the recessive one. It is also called CSMA-BA (bitwise arbitration).

2.6 Communication Paradigms

There are 2 basic communication paradigms governing the information exchange between two or more network entities [Zur14]. The first approach is built upon the cooperation of actions or functions into which more complex process can be decomposed. This paradigm is called client-server, where the responsibility to interpret information lies with the sender. The service or data providing entity is called server, and the service requesting entity is called client. The server becomes active only when its services are requested by the client, hence this paradigm suits well for state-based traffic handled in some scheduled manner.

The other approach concentrates on exchanged data rather than actions and the responsibility of its interpretation lies with the receiver. The publisher-subscriber and producer-consumer paradigms follows the data-oriented approach. In the publisher-subscriber paradigm, the information is produced by the publisher and mulicasts on the network to be listened by the subscribers. The producer-consumer model is similar to the publisher-subscriber and only differs in broadcast communication of information. There are two variants of publisher-subscriber models based on how the information exchange is initiated – pull-type, the publishing action is triggered by a centralized publishing manager,

	Client-Server	**Producer-Consumer**	**Publisher-Subscriber**
Communication relation	*Peer-to-peer*	*Broadcast*	*Multicast*
Communication type	*Connection-oriented*	*Connection-less*	*Connection-less*
Communication service	*Confirmed,unconfirmed, acknowledged*	*Unconfirmed, acknowledged*	*Unconfirmed, acknowledged*
MAC type	*Mono-master(polling, centralized TDMA), multimaster (CSMA, TDMA or Token Passing)*	*Multimaster (TDMA, centralized polling or random access)*	*Multimaster (TDMA, centralized polling or random access)*
Application class	*Parameter transfer, cyclic communication*	*Event notification, alarms, error, synchronization*	*State changes, event-oriented signal sources (eg. switches)*

Table 2.1: Properties of Communication Paradigms

and push-type, publishers become active themselves without centralized manager triggered by a timer or an event. Interestingly, in order for the subscription of the subscribers to correct communication group or multicast group the client-server-type communication is used.

The properties of these three communication paradigms are summarized in Table 2.1 [TMV95].

3 Industrial Networks and Determinism

The technical selection of networks for a particular application revolves around evaluating and balancing quality of service (QoS) parameters. Two parameters which are evaluated to find the balance between network components competing for limited bandwidth and time to deliver information between end components are network average speed and determinism. Network speed is a function of network access time and bit transfer rate. On the other hand, determinism is a measure of the ability to communicate data consistently from end to end within a

guaranteed time. The MAC component of network protocols defines the mechanism for delegating network bandwidth for optimized communication (eg. large packets with low determinism vs small packets with high determinism).

The basic QoS measures of industrial networks incorporates the speed and bandwidth of a network (i.e. how much data can be transmitted in a time interval), the delay and jitter associated with data transmission (time for a message to reach its destination and repeatability of this time), and the reliability and security of the network infrastructure.

The bandwidth of an industrial network is the number of bits that can be transmitted per second. The Ethernet-based industrial networks support data rates of 100 Mb/s or 1 Gb/s. The speed is the inverse of the data rate, thus the time to transmit 1 bit of data over the network, T_{bit} = 10 ns for 100 Mb/s Ethernet. The transmission time for a message on the network can be computed from the network's data rate, the message size and the distance between two nodes. It is considered a deterministic time in a network system. The transmission time (T_{tx}) can be written as the sum of the frame time and the propagation time:

$$T_{tx} = T_{frame} + T_{prop}.$$

Where, T_{frame} is the time required to send the packet across the network, and T_{prop} is the time for a message to propagate between any two devices.

The typical transmission speed in a communication medium is 2 x 10^8 m/s which means the propagation time T_{prop} is negligible, for example, T_{prop} = 67.2 μs for 2500 m Ethernet. The frame time (T_{frame}) depends on the size (in bytes) of data/message (N_{data}), the overhead (N_{ovhd}), padding used to meet minimum frame size requirement (N_{pad}), and the bit time (T_{bit}). Some protocols need extra bytes based on the bit-stuffing mechanism they use (N_{stuff}). The frame time can be expressed as:

$$T_{frame} = [N_{data} + N_{ovhd} + N_{pad} + N_{stuff}] \times 8 \times T_{bit}.$$

A network's time delay is defined as the total time between the sampled or computed data being available at source node and it being received and decoded at the destination node. The jitter is the variability in the delay. Many techniques have been developed to handle constant time delays however large variability in time delays is difficult to compensate for. The total time delay (T_{delay}) depends on the

preprocessing time taken at the source node for data encapsulation and encoding (T_{pre}), waiting time of the node when network is busy (T_{wait}), transmission time to send data across the network (T_{tx}) and the postprocessing time of the received data at the destination node for data decoding and postprocessing (T_{post}). T_{wait} is a function of the MAC mechanism of the protocol and can be computed based on network traffic, how many nodes are there, the relative priority of these nodes and the messages they are sending, and how much data they send. T_{pre} and T_{post} depend on the device and can be major sources of delay and jitter in a network. In an equation form, the total time delay is:

$$T_{delay} = T_{pre} + T_{wait} + T_{tx} + T_{post}.$$

The reliability of data transmission medium in a network gets affected by electromagnetic interference resulting in data corruption. To increase the reliability handshaking mechanism can be used. Acknowledgment messages (ACK) are sent between the devices to confirm the data delivery. If no ACK is received, the data is resent. However, the handshaking techniques increases the required overhead and thus decreasing the overall effective bandwidth.

Security of networked systems is another concern as the networks and operating systems are vulnerable to Internet-based attacks and viruses. Most industrial fieldbuses were not designed to be highly secure and relied on the principle of "security by obscurity" instead of authentication or encryption techniques. The intent of incorporating security in the network is usually to prevent misuse of process data than counteracting network attacks. Firewalls are installed to prevent unknown traffic from entering the network and for secure encrypted transmission virtual private network (VPN) is used. In the recent years, there has been vast progress in development of intrusion detection/prevention systems (IDS/IPS) to handle increasingly complex attacks [DNVHC05, ZJS11, McM17, BG16].

4 Industrial Communication Protocols

The first implementation of the full ISO/OSI seven-layer stack was manufacturing automation protocol (MAP) developed as a framework for the comprehensive control of industrial processes covering all automation levels. However, its complexity made implementations costly and unjustifiable for general-purpose.

Learned from the failure of MAP, further automation protocols stack was reduced and layers were combined based on the domain requirements for simplicity [GH+13] (Figure 4.1). The Internet is governed by protocols based on fully functional TCP/IP stack consisting only physical, network, transport and application layers. The IEC 61158 fieldbus standard reduced model (EPA) consists of only three layers — physical, data-link and application. Many fieldbuses are single-segment networks with limited where routing functionality and end-to-end control is not necessary. Thus, network and transport layer are removed. Also, fieldbuses were not designed for sophisticated tasks hence session and presentation layers are also not needed. However, when the layer 3 and layer 4 functions are needed they can be placed either in layer 2 or layer 7. Furthermore, layer 7 always covers layer 5 and 6 functionality.

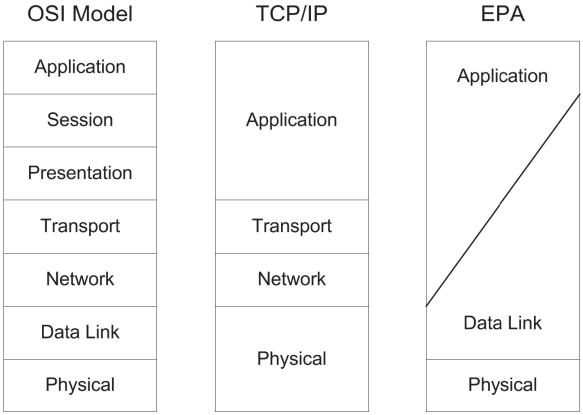


Figure 4.1: Reduced ISO/OSI stack comparison [GH+13].

As discussed earlier, Ethernet has penetrated the automation network. However, the office Ethernet didn't support deterministic capabilities hence couldn't be used for lower levels of automation pyramid. Collisions can occur on the network, and messages must be retransmitted after random amounts of time. To address this inherent nondeterminism, many different flavors of Ethernet were proposed for use in industrial automation. An effective solution in recent years has been the utilization of switches to manage the Ethernet bandwidth utilizing TDMA approach among time-critical nodes giving rise to switched Ethernet.

Switching technology does eliminate collisions, but delays inside the switches and lost packets under heavy load conditions are unavoidable also with switches. The hard real-time requirements of drive controls can't be made through these industrial Ethernet solutions. This led to development of Real Time Ethernet (RTE) standard IEC 61784.

For different application domains there are different RT performance requirements which require different implementations to achieve determinism. RTE implementations are all based on the TCP/IP model (Figure 4.2) and can be classified based on transmission time [Dec05] as follows:

- A low-speed class for human control with transmission time around 100 ms. This timing requirement is typical for the case of humans involved in the system observation, for engineering and for process monitoring. This requirement may be fulfilled with the use of Ethernet cabling and TCP and UDP for non-RT communications. This approach is called 'on top of IP' where the application layer is responsible for scheduling communication to meet the requirements. It is possible to communicate over network boundaries transparently. However, such communication introduce non-deterministic delays and the scheduling device must be equipped with adequate resources. The industrial protocols Modbus/TCP and EtherNet/IP are based on this RTE implementation.

- In the second class, for process control, the transmission time requirement is below 10 ms. This is a requirement for most tooling machine control system like PLCs. To reach this timing behavior, modification of the TCP/IP stack may be done only at the application level to use standard data packets and the transport level may be modified to use custom ethertypes for real-time communications. This approach is 'on top of Ethernet' where custom ethertypes are defined in the Ethernet frame alongside standard types such as IP. The network components and connected devices must have the knowledge of the custom protocols. Often the custom ethertypes will be given dedicated bandwidth or priority within the network. Ethernet Powerlink is the widely popular protocol within this approach-based protocols.

- The last and most demanding class is imposed by motion controls requiring a cycle time less than 1 ms with jitter not more than 1 μs. This can

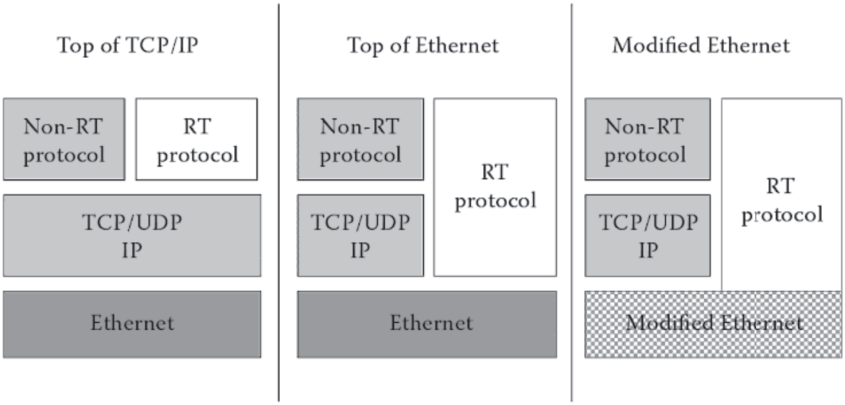


Figure 4.2: RTE implementations.

only be reached when the Ethernet data-link layer is may be modified to apply mechanisms and infrastructure that allow for real-time communication. This approach is called 'modified Ethernet' which enables non-standard topologies such as rings or buses to be implemented. To enable these topologies, the switching functionality is integrated inside the field device. The modifications are mandatory for all devices inside the RT segment but allow non-RTE traffic to be transmitted without modifications. Certain variants of PROFINET (Ethertype 0x8892), EtherCAT (Ethertype 0x88A4) and SERCOS (Ethertype 0x88CD) protocol types follow this approach to provide 1 ms transmission time requiring customized hardware.

5 Summary

This report outlined how industrial communication paradigms differ w.r.t. various communication fundamentals. The important aspect of Quality of Service (QoS) and related parameters were discussed in brief. At the end, we looked

at why and how OSI/ISO model is reduced for automation system requirements, and the approaches based on the TCP/IP model for Real Time Ethernet implementations.

Bibliography

[BG16] Anna L Buczak and Erhan Guven. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18(2):1153–1176, 2016.

[Dec05] J-D Decotignie. Ethernet-based real-time and industrial communications. *Proceedings of the IEEE*, 93(6):1102–1117, 2005.

[DNVHC05] Dacfey Dzung, Martin Naedele, Thomas P Von Hoff, and Mario Crevatin. Security for industrial communication systems. *Proceedings of the IEEE*, 93(6):1152–1177, 2005.

[DZ83] John D Day and Hubert Zimmermann. The OSI reference model. *Proceedings of the IEEE*, 71(12):1334–1340, 1983.

[GH+13] Brendan Galloway, Gerhard P Hancke, et al. Introduction to industrial control networks. *IEEE Communications Surveys and Tutorials*, 15(2):860–880, 2013.

[McM17] David McMillen. Security attacks on industrial control systems. Technical report, Technical Report. IBM, 2017.

[Shi12] Aaron Shi. OSI model data flow. `http://aaronshi.blogspot.de/2012/11/data-link-layer-add-both-header-and.html`, 2012. [Online; accessed 01-March-2018].

[SSKD11] Thilo Sauter, Stefan Soucek, Wolfgang Kastner, and Dietmar Dietrich. The evolution of factory and building automation. *IEEE Industrial Electronics Magazine*, 5(3):35–48, 2011.

[Tho05] J-P Thomesse. Fieldbus technology in industrial automation. *Proceedings of the IEEE*, 93(6):1073–1101, 2005.

[TMV95] J-P Thomesse, Zoubir Mammeri, and L Vega. Time in distributed systems cooperation and communication models. In *Distributed Computing Systems, 1995., Proceedings of the Fifth IEEE Computer Society Workshop on Future Trends of*, pages 41–49. IEEE, 1995.

[Zim80] Hubert Zimmermann. OSI reference model – the ISO model of architecture for open systems interconnection. *IEEE Transactions on communications*, 28(4):425–432, 1980.

[ZJS11] Bonnie Zhu, Anthony Joseph, and Shankar Sastry. A taxonomy of cyber attacks on SCADA systems. In *Internet of things (iThings/CPSCom), 2011 international conference on and 4th international conference on cyber, physical and social computing*, pages 380–388. IEEE, 2011.

[Zur14] Richard Zurawski. *Industrial communication technology handbook*. CRC Press, 2014.

Phase Detection in Medical Context: Overview and Outlook

Patrick Philipp

Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT), Germany
p.philipp@kit.edu

Technical Report IES-2017-05

Abstract: To provide assistance functions, e.g. in context of surgical interventions, the use of a phase detection plays an important role. For instance, by assessing the progress of an on-going surgery, a tailored (i.e. context sensitive) decision support for medical practitioners can be carried out. The optimization of a workflow, e.g. by comparing recorded data with a predefined target model, is another application example. Subsequently, a phase detection provides opportunities to prevent errors, injuries, negligence or malpractices in the medical context. In this work, an overview of notable model approaches for a phase detection in medical context is presented. Based on this, further suggestions for future models are proposed.

1 Introduction

In contemporary medicine, the use of advanced computer-based assistance (including both: hardware and software) becomes increasingly important [PFHB16]. E.g. the global market for medical robotics and computer-assisted surgical equipment is projected to grow to 6.8 billion dollars by 2021 (using a five-year compound annual growth rate of 11.3%) [McW17].

As part of a computer assisted surgery (CAS), assistance functions can be realized to enable a decision support for medical practitioners [KWN+15]. Thereby, a decision support opens up a field of optimization, e.g., concerning the prevention of errors, negligence, injuries or (as a result) malpractices.

In this context, a surgical phase detection plays an important role. Namely, because by assessing the progress of an on-going surgery, a tailored (i.e., context sensitive) and interactive decision support during an intervention can be enabled. In doing so, there is not only a passive dissemination (e.g. distribution via print media) of support (e.g. medical guidelines) – which has only little effect on the actual practitioners behavior [FL92, SGM+11]. Moreover, a tailored decision support allows for filtering physically available information in order to be operationally effective by preventing an informational overload of medical practitioners [JL83, KSF+13, LDC+13].

The optimization of a workflow, e.g. by comparing recorded data with a predefined target model, is another application example of a phase detection. This is especially relevant as surgeries have been identified as an important source of improvement to the hospital efficiency [Her03].

In this work, an overview of notable model approaches for a phase detection in medical context are presented. Based on this, further suggestions for elaborating future models are proposed. This contribution is structured as follows: first in Section 2, the constituents of the phase detection problem are elaborated. In Section 3 notable modeling approaches are presented and summarized in Section 4. Section 5 focuses on future implications for models concerning a phase detection and, finally, a conclusion is drawn in Section 6 on page 78.

2 Constituents of the Problem

Figure 2.1 depicts a workflow of a surgery using an UML activity diagram [OMG11]. This formalism is also used in previous research concerning the modeling of medical workflows [PFHB15a, PFHB15b, Phi16, PFHB16, PFB17].

The start of the activity “Surgery” is symbolized by a solid circle (inital node), whereas the end of the activity is given by a double circle (activity final). Performed actions are shown as rounded rectangles, while the control flow is represented by arrows (directed edges). In this example, the actions are performed sequentially – i.e. one after another.

The notation elements “...”, the red rectangle as well as the red arrows are added as visual aids – they are not part of the UML specification [OMG11]. Thereby

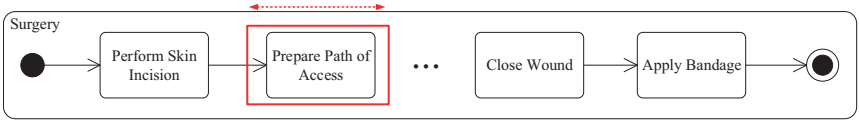


Figure 2.1: Figure shows a pre-modeled workflow of a surgery. Goal of a phase detection is to match a pre-modeled workflow with the current conditions in the operating room.

"..." symbolize a sequences of arbitrary actions which are not shown in the Figure for simplification. The red rectangle highlights the current predicted action, which changes over time.

Goal of a phase detection is to match online a pre-modeled workflow with the current situation in the operating room. It means in effect (cf. Figure 2.1) that we want to know, e.g., is the surgeon currently performing the skin incision or is he already preparing the path of access.

To reach this goal, feature values x are made available by sensors in the operating room. Such values are e.g. "number of persons at the table", "number of used instruments" and so forth. These values are bundled by a feature vector $\boldsymbol{x}$ which is defined as a column vector whose elements are these feature values x – i.e.

$$\boldsymbol{x} = (x^1, \ldots, x^D)^T \, .$$

Whereby D represents the number of features, i.e. the dimension of the resulting feature space as well as of the corresponding feature vector. To save space, vector $\boldsymbol{x}$ is shown horizontal (i.e. transposed, indicated by T).

In each time step a new feature vector $\boldsymbol{x}$ with index t is observed by the technical system in the operating room:

$$\boldsymbol{x}_t = (x_t^1, \ldots, x_t^D)^T \, .$$

As a consequence, there is an observation sequence of feature vectors

$$\boldsymbol{x}_{1:t} = \boldsymbol{x}_1, \ldots, \boldsymbol{x}_t \, .$$

The observation sequence $\boldsymbol{x}_{1:t}$ is available to the technical system, whereas the current phase is not. I.e. there is a gap between the observation sequence $\boldsymbol{x}_{1:t}$

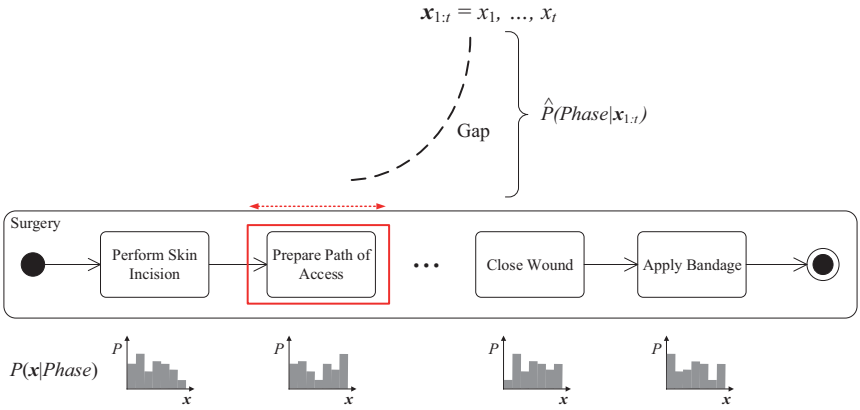


Figure 2.2: Figures depicts the gap between a sequence of feature vectors and the surgical phase. The latter is hidden to the technical system (w.l.o.g.). To bridge the gap between them, a probability distribution P can be used to specify how likely it is to emit a certain feature vector being at a specific phase (i.e. $P(\boldsymbol{x}|Phase)$). A Hidden Markov Model (HMM) can be used to infer the probability of being in a phase given the observation sequence.

and the surgical phase which is not directly visible (i.e. hidden) to the technical system (w.l.o.g. cf. Figure 2.2).

In traditional approaches like Hidden Markov Models (HMM), a probability distribution P is used to specify how likely it is to emit a certain feature vector being at a specific phase (cf. Figure 2.2 bottom row).

Using further parameters of the HMM (cf. Section 3.3), vice versa, the probability of being in a Phase given the observation sequence can be inferred. As a consequence the gap between the observation sequences and a specific phase is bridged. Further model characteristic details are elaborated in the following sections.

3 Modeling Approaches

In the following section notable modeling approaches for a medical phase detection are elaborated. Thereby, the different characteristics of the models are illustrated.

3.1 Random Forest (RF)

A RF is a model which derives its decision from a set of classifiers and therefore belongs to the so called ensemble methods [Rok10, Zho12]. A RF predicts the class of a feature vector $f(\boldsymbol{x})$ by a weighted sum [Bre01]

$$f(\boldsymbol{x}) = \sum_{n=1}^{N} \alpha_n t_n(\boldsymbol{x}) \,.$$

Thereby a set of decision trees (i.e. a forest) is used:

$$\{t_n(\boldsymbol{x}) : n = 1, \ldots, N\} \,.$$

These classifications results are equally weighted, i.e. $\alpha_n = (N)^{-1}$. Conceptually, a RF utilizes randomization in two ways:

Firstly, the decision trees $t_n(\boldsymbol{x})$ are trained by randomly sampling the training set $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ with known class memberships $\omega(\boldsymbol{x}_i)$ for $i = 1, \ldots, M$. The idea behind this is, to lower the variance by averaging over a large number of trees which are highly uncorrelated (bootstrap aggregation aka bagging).

Secondly, during the training of a decision tree, for each node only a random subset of feature values are considered for splitting the tree. I.e. the size of the considered feature vector $\boldsymbol{x}$ in each split is $d < D$. Based on a quality metric (e.g. entropy or genie) and under consideration of d feature values, the split with the highest quality is calculated. Stopping criteria of the training algorithm of a tree is, e.g., that a minimum size of a leaf is reached. I.e. the split of a set of feature vectors would result in subsets that are too small.

For classification, each decision tree $t_n(\boldsymbol{x})$ of the RF receives the same feature vector $\boldsymbol{x}$ and maps it to a class $\omega \in \Omega$. The corresponding a-posteriori probability is given by:

$$\hat{P}(\omega|\boldsymbol{x}) = \frac{1}{N} \sum_{n=1}^{N} \left[t_n(\boldsymbol{x}) = \omega\right] \,. \tag{3.1}$$

Please note that $[\cdot]$ represents a predicate mapping, i.e. $[\cdot]$ has a value of 1 iff the corresponding expression evaluates to true.

The predicted class $\hat{\omega}$ is then given by [Bre01]:

$$\hat{w} = \arg\max_{\omega \in \Omega} \hat{P}(\omega|\boldsymbol{x}) \,. \tag{3.2}$$

The use of the RF model for a medical phase detection by the application example of a cholecystectomy (removal of the gallbladder) is elaborated in [SOP+14]. 7 phases are considered in context of this surgery. The training set comprises 9 surgical interventions. Nominal features are the use of eight specific instruments (yes/no), the state of the surgical light (on/off) as well as the state of the room lights (on/off). The authors claim that an RF is suitable, because it is highly applicable for multi-class problems and it is able to predict phases in a-typical order. The latter is a consequence of the fact that a RF does only consider a single feature vector for classification (cf. Equations (3.1) and (3.2)). That means in effect, that a feature vector is classified without taking the order of phases into account.

The trained model achieved an accuracy of around 69% for the 7 classes using cross-validation with a leave-one-out iterator. Classes with similar characteristics (which differ only in the sequence of occurrence) can not be distinguished. In [SPG+16] the authors also get comparable results for a simplified 7-phase hip replacement surgery with a high confusion for phases which are highly discriminative to the order of phases.

This is a prototypical behavior of a classifier that does not take the sequence of feature vectors into account. Therefore this model is, out of the box, not suited regarding the constituents of the problem. Nevertheless, to reduce the conceptional drawback, such a model can be embedded into a sequential model. One simple form of such a model is, e.g., a deterministic automaton.

In [KSW+16] this concept (cf. Figure 3.1) is used on two application examples: Firstly, a pancreatic resection (removal of the pancreas) comprising 12 classes and a training set of 11 surgeries. Secondly a adrenalectomy (removal of the adrenal glands) with 9 classes and a training set of 5 surgeries. Considered features are the use of a specific instrument, the action performed and the anatomical structure.

The use of this sequential framework (cf. Figure 3.1) is like shifting a window over a workflow (cf. Figure 2.1) comprising the current phase, and additionally,

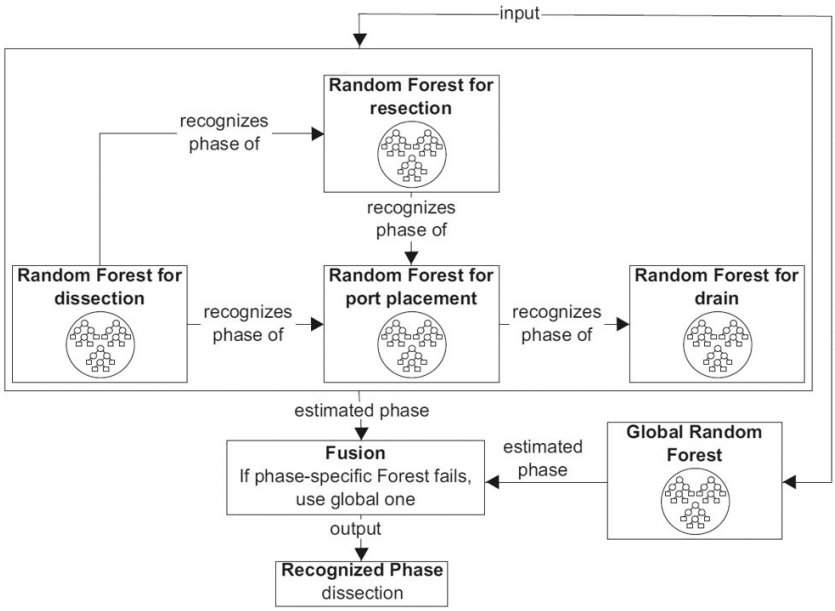


Figure 3.1: Figures depicts the utilization of a random forest (RF) embedded into a simple sequential model. A RF is used to predict the current and possible next phases. (e.g. dissection with possible next phases resection and port placement). If a next phase is predicated, a transition is made. As a result, another RF is used to discriminate between the new current phase and possible next phases. If a transition is made by mistake, the model tends to get lost. Therefore a global RF is used to discriminate between all possible phases in order to be able to reset the sequential model. Modified from [KSW+16].

possible next phases. A window-specific classifier then discriminates between the phases inside the window [Phi16, Phi17]. The window is shifted forward as soon as a next phase is predicted.

The modeling approach in [KSW+16] achieved an accuracy around 70%. Drawback is the use of deterministic transitions, which makes it necessary to implement strategies to reset the sequential model (cf. Figure 3.1).

3.2 Stochastic Petri Net (SPN)

There are a whole range of reasons for considering Petri Nets [Pet62] as a modeling tool for dynamic aspects of a process [vdA96, vdA98]. With respect to the application example, Petri Nets can be of use because of their formal semantics and the abundance of analysis techniques.

The net structure [Rei13b] of a Petri Net is given by the tupel

$$NST = (\mathcal{P}, \mathcal{T}, \mathcal{F}) ,$$

where $\mathcal{P}$ is the set of places and $\mathcal{T}$ is the set of transitions

$$\mathcal{P} = \{p_i : i = 1, \ldots, |\mathcal{P}|\} ,$$
$$\mathcal{T} = \{t_i : i = 1, \ldots, |\mathcal{T}|\} ,$$

so that

$$\mathcal{P} \cap \mathcal{T} = \emptyset .$$

The flow relation $\mathcal{F}$ reflects the connection of places and transitions (and vice versa):

$$\mathcal{F} \subseteq (\mathcal{P} \times \mathcal{T}) \cup (\mathcal{T} \times \mathcal{P}) .$$

Consequently, the net structure NST of a Petri Net is a directed bipartite graph (see Figure 3.2). To model the dynamic behavior of the system, so called "tokens" are introduced (black dot in Figure 3.2). The distribution of tokens on the set of places represents the state of the Petri Net. It is also called marking [Rei13b].

A Stochastic Petri Net (SPN) is defined as a tuple

$$SPN = (NST, \Lambda) ,$$

whereby $\Lambda = \{\lambda_i : i = 1, \ldots, |\mathcal{T}|\}$ is the set of shifting rates λ_i which are assigned to transitions t_i. These shifting rates are distributed exponentially [Mol81].

Figure 3.2 depicts a SPN on the left. Circles represent places p_i and squares represent transitions t_i. The state of the SPN is given by the current distribution of tokens (marking). In the reachability graph the states are represented by vectors, whereby each entry i represent the number of tokens at a place p_i. That means,

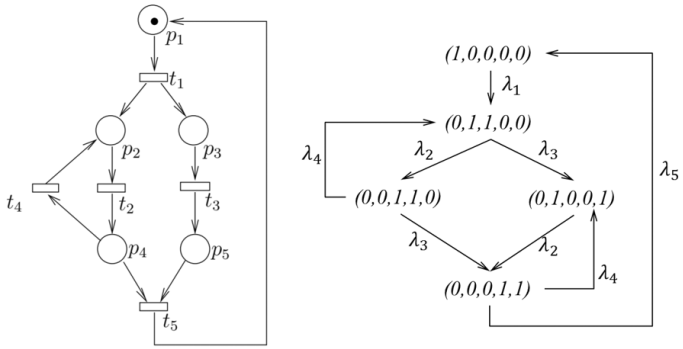


Figure 3.2: Figure depicts a SPN on the left. Circles represent places p_i and squares represent transitions t_i. The state of the SPN is given by the current distribution of tokens (marking). In the reachability graph the states are represented by vectors, whereby each entry i represent the number of tokens at a place p_i. The Figure visualizes that a SPN represents a Markov Process whereby, inter alia, the firing rates λ_i of states Z_i of the reachability graph corresponds to the transition probabilities a_{ij} of a (discrete) Markov Process. Modified from [B$^+$02].

the state $(1, 0, 0, 0, 0)$ of the reachability graph represents the state of the SPN depicted on the left. Figure 3.2 visualizes the fact that a SPN actually represents a Markov Process. I.e. the states of the reachability graph of the SPN correspond to the states of a Markov Process and the firing rates λ_i of states Z_i of the reachability graph correspond to the transition probabilities a_{ij} of a (discrete) Markov Process. That is, inter alia, because the firing rates are exponentially distributed (cf. memorylessness) and therefore the Markov Property is satisfied [B$^+$02].

The use of Markov Models with observable states is suitable in cases where the states of the modeled system are directly accessible. E.g. in [PFHB15a] a decision support system for the diagnosis of two complex cancerous diseases is modeled using Petri Nets. The involved dialog system allows for a direct access to the necessary features. Clearly, this model can also be used as sequential model to embed a model like a RF instead of a deterministic automaton (cf. Section 3.1). Nevertheless, with respect to the constituents of the problem, this model has to be extended, e.g. by introducing hidden states.

3.3 Hidden Markov Model (HMM)

A Hidden Markov Model (HMM) [RJ86] is a Markov Model using a Markov Process with unobserved (i.e. hidden) states. That means, in contrast to Markov Processes, where the state of the model at a time step t is known for certain (i.e. directly observable), in a HMM an observer can only access emissions which are generated by a hidden (i.e. not observable) state. To enable this, state transition probabilities and furthermore emission probabilities are part of a HMM.

Formally, a (discrete) HMM is defined as a 5-tuple

$$\lambda = (S, V, A, B, \pi)\,,$$

whereby $S = \{s_1, \ldots, s_n\}$ is the set of hidden states and $V = \{v_1, \ldots, v_m\}$ is the set of emissions for each hidden state (i.e. output vocabulary). Moreover, there is a transition matrix $A \in \mathbb{R}^{n \times n}$ encoding the transition probabilities from a current state s_i to a next state s_j by the matrix entry (a_{ij}). Matrix $B \in \mathbb{R}^{n \times m}$ encodes the emission probabilities of a state s_i by the corresponding row entries $(a_{ij}) : j = 1, \ldots, m$. Finally, $\pi \in \mathbb{R}^n$ encodes the initial state probabilities, i.e. the probability that a state is the starting state.

The use of a HMM for a medical phase detection has been elaborated in [BPFN08] by the application example of a cholecystectomy (removal of the gall-bladder). 14 phases are considered in context of this surgery. The training set comprises 12 surgical interventions. Nominal features are the use of 17 specific instruments (yes/no). The authors present a 14-state HMM and a merged HMM. In the latter, similar states are transformed into one single state. The models achieved an accuracy of around 86% and 93% respectively, using a complete cross-validation.

In [PBF+08] a cholecystectomy comprising 14 phases is considered, too. The training set comprises 11 surgical interventions. Nominal features are the use of 18 specific instruments (yes/no) including the state of an optical device (inserted/not inserted). The best accuracy achieved using a complete cross-validation is around 92% – but for computing this value, a tolerance of 5 seconds before and after the ground truth definition of a phase is set. Additionally. the authors state that phases with very short durations are poorly recognized which is a result of the inertia of the model.

A HMM can be seen as a common sequential model considering the constituents of the problem (cf. Section 2). Indeed, it can be a suitable model for problems where data is not independent and identically distributed, i.e. a feature vector $\boldsymbol{x}$ does depend on the feature vectors seen before. Nevertheless, the structure of this model is not well suited to include expert knowledge about emission probabilities. That is because, inter alia, the corresponding probability distribution can not be factorized and therefore grows exponentially with the number of possible emissions. To overcome this drawback, the generalization of a HMM, namely a Dynamic Bayesian Network can be used.

3.4 Dynamic Bayesian Network (DBN)

A Bayesian Network (BN) is a probabilistic graphical model (PGM), combining graph theoretic approaches with approaches of probability theory. Consequently, a BN over random variables $X^{0:N} := X^0, \ldots, X^N$ is given by a pair

$$B = (G, P)\,.$$

Whereby G corresponds to a directed, acyclic graph

$$G = (V, E)\,,$$

and

$$P(X^{0:N}) = \prod_{n=0}^{N} P(X^n|\mathrm{Pa}(X^n))\,,$$

corresponds to a joint probability distribution [KF09].

Graph G is used to define dependencies between random variables $X^{0:N}$. It is also known as the structure of the BN. The vertex set V represents the set of random variables, while a directed edge $V_i \rightarrow V_j$ of the set of edges E represents a direct dependency between two variables. A missing edge symbolizes the independence of these two variables.

The joint probability distribution is given by the product of all conditional probability distributions associated with the vertices of G. It is also known as the parameters of the BN. Here, $\mathrm{Pa}(X^n)$ denotes the set of parents of a random variable X^n. Graphically, this corresponds to vertices having a directed edge pointing to

X^n's vertex. Please note, if $\text{Pa}(X^n) = \emptyset$, a random variable X^n is a root node of the BN, and $P(X^n|\,\emptyset) = P(X^n)$ gives the a-priori probability.

A Dynamic Bayesian Network (DBN) is an extension of a BN, also taking the temporal dependencies of variables into account [Mur02]. A DBN is given by a pair

$$\text{DBN} = (B_0, B_{\rightarrow})\,,$$

where the BN B_0 uses $P(X_0^{0:N})$ to specify the a-priori probability distribution over random variables $X^{0:N}$ in a time step with index 0.

Furthermore, $B_{\rightarrow}$ specifies the conditional probability distribution over discrete time steps t by using

$$P(X_t^{0:N}|X_{t-1}^{0:N}) = \prod_{n=0}^{N} P(X_t^n|\text{Pa}(X_t^n))\,.$$

Thereby $\text{Pa}(X_t^n)$ denotes the set of X_t^n's parents in the corresponding graph. The parents can be in the same time slice (e.g. representing instantaneous causation) or the previous one (i.e., we assume the model to be fist-order Markov). In the latter case, arcs point to time slices with ascending index, reflecting the causal flow of time [Mur02].

Figure 3.3 depicts on the left a DBN structure that represents a HMM. Thereby the root node represents a surgical phase and the child node represents the emission of feature vectors $\boldsymbol{x}_t$ given a phase, i.e. $P(\boldsymbol{x}_t|\mathit{Phase}_t)$. By introducing a conditional independence between the feature values (cf. Figure 3.3, right side), a naive DBN can be constructed. The corresponding probability distribution of the children is given by $\prod_{n=1}^{N} P(\boldsymbol{x}_t^n|\mathit{Phase}_t)$. It can easily be seen that in the first case the number of parameters grows exponentially, whereas in the second case the number of parameters grows linearly. For example, let's assume that there are 7 distinct phases to predict and there is a feature vector $\boldsymbol{x}_t$ with dimension $D = 3$ having discrete feature values with 4 different characteristics. Then, the corresponding conditional probability $P(\boldsymbol{x}_t|\mathit{Phase}_t)$ of a HMM is given by $(4 \cdot 4 \cdot 4) \cdot 7 = 448$ parameters from which 441 have to specified because the probabilities some up to 1. But $\prod_{n=1}^{N} P(\boldsymbol{x}_t^n|\mathit{Phase}_t)$ is given by $(4 + 4 + 4) \cdot 7 = 84$ parameters from which 63 parameters have to specified. Further details can also be found in [PFB17].

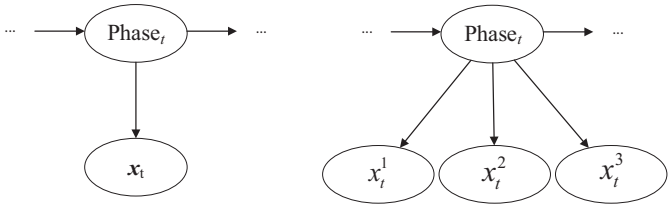


Figure 3.3: Figure shows on the left a HMM represented by a DBN structure. There is a root node representing the surgical phases and a child node representing the emissions given a phase. The parameters needed to specify a conditional probability distribution of this child node grow exponentially with the number of used features, i.e. with the size of the feature vector $\boldsymbol{x}_t$. On the right side of the figure a DBN structure is shown which takes advantage of conditional independence of feature values x_n^t to reduce the growing of parameters: in the depicted case the number of parameters linearly grows with the size of $\boldsymbol{x}_t$.

This reduction of parameters can be useful in case only a small amount of training data is available [KF09]. Furthermore, the special structure of the network supports expert-based parametrization and therefore to incorporate the knowledge of experts into the models – which also can be useful when dealing with small amount of training data and in context of translation and/or fusion mechanisms of workflow models [PFHB15b, PFHB16, PFB17].

In [PSG+16] a DBN model is elborated by using the application example of a simplified total hip replacement. 7 phases are considered in context of this surgery and the training set comprises 12 surgical interventions. Features are the positioning of the sergeants (position A / position B), the number of used instruments (0/1/2) as well as the number of persons at the operating table (0/1/2/3/4). Using a complete cross-validation, the model achieves an accuracy around 85%. Concerning an earlier publication using RF [SPG+16] the sequence modeling greatly reduces confusion of time depended feature values.

4 Overview

The model approaches discussed in Section 3 are briefly summarized in Table 4.1. Thereby a Random forest (RF) shows prototypical behavior of a model

which omits the sequence of feature vectors $\boldsymbol{x}_t$ during classification. Although this drawback can be addressed, e.g. by embedding such a classifier into a sequential model, a RF can generally categorized as non-sequential. Subsequently RF do no allow to model hidden states, which can be necessary to represent the fact that an observer can not directly access the state of a modeled system (cf. Section 2). Furthermore a RF is a discriminative model and allows for an online classification. The latter is necessary to be able to provide a phase detection on the fly (cf. Section 2). The model allows for a data-driven training but lacks of a suitable interface to incorporate expert knowledge which can be especially of use if only a small amount of training data is available.

A Stochastic Petri Net (SPN) represents a Markov Process (cf. Section 3.2). This model is suited for representing sequential dependencies for systems in which internal states are visible to an observer (e.g. Dialogue Systems). Furthermore, SPN is an generative approach allowing for an online classification. A data-driven training is possible, e.g. by using genetic algorithms. Finally, a knowledge-based parametrization of a SPN is possible, too – in practice the size of the model can cause limitations. In such a case, an extension of the model by using additional concepts, e.g. Coloured Petri Nets, can become necessary.

A Hidden Markov Model (HMM) can be seen as a common and well researched sequential model. A HMM allows for a direct modeling of hidden states which suits the considered application example (cf. Section 2). It is an generative approach and is able to classify online. Nevertheless, the structure of this model is not well suited to incorporate expert knowledge about the emission probabilities. That is because, inter alia, the corresponding probability distribution can not be factorized and therefore grows exponentially with the number of possible emissions.

To overcome this drawback, the generalization of a HMM, namely a Dynamic Bayesian Network (DBN) can be used. It extends the model by the ability to express the state space in a factored form and not only as a single random variable. This allows for the reduction of parameters and consequently facilitates the improvement of modularity and interpretability. This opens up a practicable way to incorporate expert knowledge into a DBN. Furthermore, concerning Kalman Filter Models, a DBN allows for arbitrary probability distributions (not only for

	RF	SPN	HMM	DBN
Sequential Model	□	■	■	■
Hidden States	□	□	■	■
Generative Approach	□	■	■	■
Online Classification	■	■	■	■
Data-driven Training	■	■	■	■
Knowledge-based Parametrization	□	■	□	■

Table 4.1: The Table briefly compares different modeling approaches discussed in this work. The symbols ■ / □ are used to specify if a property is present / not present.

unimodal linear-Gaussians). A DBN is a promising approach because it combines a reasonable tradeoff between expressiveness and complexity, and includes probabilistic models that have proved to be successful in practice (e.g. HMM).

5 Outlook

Recently, the use of Artificial Neural Networks is elaborated. E.g. a Convolutional Neural Network (CNN) for a medical phase detection is used in [LZL+16] by the application example of a resuscitation with only 3 phases. The training set comprises 20 workflows. Features are extracted from depth image recordings and from sound recordings. Furthermore, the model is embedded into a simple sequential model represented by thresholds which is similar to a finite automaton. Nevertheless, the model achieved an accuracy of 80%.

In [TSM+17] a cholecystectomy comprising 7 phases is considered. The training set comprises 80 surgical interventions. The authors use a CNN to extract features from video recordings of endoscope. The model is embedded into a HMM, which achieved an accuracy of 82%. Finally, in [LZZ+17] the authors elaborated the application example of a resuscitation with 35 classes and a training set of 42 video recordings. Also a CNN is used for feature extraction – in this case, the CNN is combined with a recurrent neural network using long-short term memory (LSTM). Features are extracted from depth image records, sound recordings and via a passive RFID (radio-frequency identification) system which tracks the medical instruments. The model achieves an accuracy of around 94%.

In view of these results, a further study of Artificial Neural Networks with respect to the constituents of the problem (cf. Section 2) can be considered as useful. Further elaboration is needed e.g. concerning the incorporation of expert knowledge (cf. [HML+16, PBFB17]) into the models, the challenge of a small amount of available data for training and the explainability of the models output [RSG16].

6 Conclusion

In this work, we discussed notable modeling approaches for a surgical phase detection. The models have different characteristics and are differently suited concerning the constituents of the problem. Considering a small amount of training data and a preferably expert-based modeling, Dynamic Bayesian Networks seems to be a suited and long-standing solution. Nevertheless, upcoming approaches, facilitating Artificial Neural Networks, have recently shown promising classification results in the field of surgical phase detection. To trim these models to be accessible to expert-based knowledge and be able to deal with the fact that typically only a small amount of real training data is available, could be a good starting point for approaches which enrich the current state of the art.

Bibliography

[B+02] Falko Bause et al. *Stochastic Petri nets – An introduction to the theory*. Citeseer, 2002.

[BPFN08] Tobias Blum, Nicolas Padoy, Hubertus Feußner, and Nassir Navab. Modeling and online recognition of surgical phases using hidden Markov models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 627–635. Springer, 2008.

[Bre01] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[FL92] E Field and K Lohr. *Guidelines for Clinical Practice: From Development to Use*. National Academies Press, 1992.

[Her03] C Herfarth. Lean surgery through changes in surgical work flow. *British Journal of Surgery*, 90(5):513–514, 2003.

[HML+16] Zhiting Hu, Xuezhe Ma, Zhengzhong Liu, Eduard Hovy, and Eric Xing. Harnessing deep neural networks with logic rules. *arXiv:1603.06318*, 2016.

[JL83] Joseph P Joyce and George W Lapinsky. A history and overview of the safety parameter display system concept. *IEEE Transactions on Nuclear Science*, 30(1):744–749, 1983.

[KF09] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[KSF+13] Michael Kranzfelder, Christoph Staub, Adam Fiolka, Armin Schneider, Sonja Gillen, Dirk Wilhelm, Helmut Friess, Alois Knoll, and Hubertus Feussner. Toward increased autonomy in the surgical OR: needs, requests, and expectations. *Surgical endoscopy*, 27(5):1681–1688, 2013.

[KSW+16] Darko Katić, Jürgen Schuck, Anna-Laura Wekerle, Hannes Kenngott, Beat Peter Müller-Stich, Rüdiger Dillmann, and Stefanie Speidel. Bridging the gap between formal and experience-based knowledge for context-aware laparoscopy. *International journal of computer assisted radiology and surgery*, 11(6):881–888, 2016.

[KWN+15] HG Kenngott, M Wagner, F Nickel, AL Wekerle, A Preukschas, M Apitz, T Schulte, R Rempel, P Mietkowski, F Wagner, et al. Computer-assisted abdominal surgery: new technologies. *Langenbeck's Archives of Surgery*, 400(3):273–281, 2015.

[LDC+13] Cristian A Linte, Katherine P Davenport, Kevin Cleary, Craig Peters, Kirby G Vosburgh, Nassir Navab, Pierre Jannin, Terry M Peters, David R Holmes III, Richard A Robb, et al. On mixed reality environments for minimally invasive therapy guidance: systems architecture, successes and challenges in their implementation from laboratory to clinic. *Computerized Medical Imaging and Graphics*, 37(2):83–97, 2013.

[LZL+16] Xinyu Li, Yanyi Zhang, Mengzhu Li, Shuhong Chen, Farneth R Austin, Ivan Marsic, and Randall S Burd. Online process phase detection using multimodal deep learning. In *Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE Annual*, pages 1–7. IEEE, 2016.

[LZZ+17] Xinyu Li, Yanyi Zhang, Jianyu Zhang, Shuhong Chen, Ivan Marsic, Richard A Farneth, and Randall S Burd. Concurrent activity recognition with multimodal CNN-LSTM structure. *arXiv preprint arXiv:1702.01638*, 2017.

[McW17] Andrew McWilliams. Medical robotics and computer-assisted surgery: The global market. *BCC Research*, HLC036G, 2017.

[Mol81] Michael Karl Molloy. *On the integration of delay and throughput measures in distributed processing models*. University of California, Los Angeles, 1981.

[Mur02] Kevin Patrick Murphy. *Dynamic Bayesian networks: representation, inference and learning*. PhD thesis, University of California, Berkeley, 2002.

[OMG11] OMG. OMG Unified Modeling Language(OMG UML) Superstructure Version 2.4.1, 2011.

[PBF+08] Nicolas Padoy, Tobias Blum, Hubertus Feussner, Marie-Odile Berger, and Nassir Navab. On-line recognition of surgical activity for monitoring in the operating room. In *AAAI*, pages 1718–1724, 2008.

[PBFB17] Patrick Philipp, Johannes Bleier, Yvonne Fischer, and Jürgen Beyerer. Towards a surgical phase detection using Markov logic networks. *Radermacher, Klaus (Ed.): CAOS*

2017, 17th Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery. Papers. Online resource: June 14-17, 2017, Aachen, Germany. (EPiC Series in Health Sciences 1), pp. 288-294, 2017.

[Pet62] Carl Adam Petri. *Kommunikation mit Automaten*. PhD thesis, Universität Bonn, 1962.

[PFB17] Patrick Philipp, Yvonne Fischer, and Jürgen Beyerer. Expert-based probabilistic modeling of workflows in context of surgical interventions. *CogSima 2017, IEEE Conference on Cognitive and Computational Aspects of Situation Management*, 2017.

[PFHB15a] Patrick Philipp, Yvonne Fischer, Dirk Hempel, and Jürgen Beyerer. Framework for an Interactive Assistance in Diagnostic Processes Based on the Translation of UML Activities into Petri Nets. In *Proceedings of ISHI 2015 – International Symposium on Health Informatics and Medical Systems: CSCI 2015. International Conference on Computational Science and Computational Intelligence*, pages 732–737. IEEE Conference Publishing Services, 2015.

[PFHB15b] Patrick Philipp, Yvonne Fischer, Dirk Hempel, and Jürgen Beyerer. Modeling of clinical practice guidelines for interactive assistance in diagnostic processes. In *WorldComp 2015, World Congress in Computer Science, Computer Engineering, and Applied Computing : HIMS 2015, International Conference on Health Informatics and Medical Systems, july 27-30, Las Vegas, Nevada, USA*, pages 3–9. CSREA Press, 2015.

[PFHB16] Patrick Philipp, Yvonne Fischer, Dirk Hempel, and Jürgen Beyerer. Framework for an interactive assistance in diagnostic processes based on probabilistic modeling of clinical practice guidelines. In *Emerging Trends in Applications and Infrastructures for Computational Biology, Bioinformatics, and Systems Biology*, pages 371–390. Elsevier, 2016.

[Phi16] Patrick Philipp. Framework for modeling medical guidelines based on the translation of UML activities into YAWL. Technical report, KIT Scientific Publishing, Karlsruhe, 2016.

[Phi17] Patrick Philipp. Combining YAWL and DBNs for surgical phase detection. Technical report, KIT Scientific Publishing, Karlsruhe, 2017.

[PSG+16] Patrick Philipp, Luzie Schreiter, Johannes Giehl, Yvonne Fischer, Joerg Raczkowsky, Markus Schwarz, Heinz Woern, and Jürgen Beyerer. Situation detection for an interactive assistance in surgical interventions based on dynamic Bayesian networks. *CRAS 2016, 6th Joint Workshop on New Technologies for Computer/Robot Assisted Surgery,*, 2016.

[Rei13b] Wolfgang Reisig. *Understanding Petri Nets*. Springer, Heidelberg, 2013.

[RJ86] Lawrence Rabiner and B Juang. An introduction to hidden Markov models. *ieee assp magazine*, 3(1):4–16, 1986.

[Rok10] Lior Rokach. *Pattern classification using ensemble methods*, volume 75. World Scientific, 2010.

[RSG16] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Why should I trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144. ACM, 2016.

[SGM+11] Earl Steinberg, Sheldon Greenfield, Michelle Mancher, et al. *Clinical Practice Guidelines We Can Trust*. National Academies Press, 2011.

[SOP+14] Ralf Stauder, Aslı Okur, Loïc Peter, Armin Schneider, Michael Kranzfelder, Hubertus Feussner, and Nassir Navab. Random forests for phase detection in surgical workflow analysis. In *International Conference on Information Processing in Computer-Assisted Interventions*, pages 148–157. Springer, 2014.

[SPG+16] Luzie Schreiter, Patrick Philipp, Johannes Giehl, Yvonne Fischer, Joerg Raczkowsky, Markus Schwarz, Jürgen Beyerer, and Heinz Woern. Situation detection for an interactive assistance in surgical interventions based on random forests. *in Proceedings of International Journal of Computer Assisted Radiology and Surgery (CARS),*, pages 115–116, 2016.

[TSM+17] Andru P Twinanda, Sherif Shehata, Didier Mutter, Jacques Marescaux, Michel de Mathelin, and Nicolas Padoy. Endonet: A deep architecture for recognition tasks on laparoscopic videos. *IEEE transactions on medical imaging*, 36(1):86–97, 2017.

[vdA96] Wil van der Aalst. Three good reasons for using a Petri-net-based workflow management system. In *Proceedings of the International Working Conference on Information and Process Integration in Enterprises*, pages 179–201. World Scientific, 1996.

[vdA98] Wil van der Aalst. The application of Petri nets to workflow management. *Journal of circuits, systems, and computers*, 8(01):21–66, 1998.

[Zho12] Zhi-Hua Zhou. *Ensemble methods: foundations and algorithms*. CRC press, 2012.

Deep Learning based Vehicle Detection in Aerial Imagery

Lars Sommer

Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT), Germany
lars.sommer@kit.edu

Technical Report IES-2017-06

Abstract: Detecting vehicles in aerial images is an important task for many applications like traffic monitoring or search and rescue work. In recent years, several deep learning based frameworks have been proposed for object detection. However, these detection frameworks were developed and optimized for datasets that exhibit considerably differing characteristics compared to aerial images, e.g. size of objects to detect. In this report, we demonstrate the potential of Faster R-CNN, which is one of the state-of-the-art detection frameworks, for vehicle detection in aerial images. Therefore, we systematically investigate the impact of adapting relevant parameters. Due to the small size of vehicles in aerial images, the most improvement in performance is achieved by using features of shallower layers to localize vehicles. However, these features offer less semantic and contextual information compared to features of deeper layers. This results in more false alarms due to objects with similar shapes as vehicles. To account for that, we further propose a deconvolutional module that up-samples features of deeper layers and combines these features with features of shallower layers.

1 Introduction

Vehicle detection in aerial images is an important task for many applications like traffic monitoring or search and rescue work. Conventional approaches applied to detect vehicles in aerial images are generally comprised of hand-crafted features

and a classifier within a sliding window approach [LM15, CH16, MM14]. In recent years, several authors applied convolutional neural networks (CNNs) to extract features at each sliding window position [CXLP14, KPF16]. In [CXLP14], improved results are achieved for vehicle detection in satellite images by applying convolutional features instead of hand-crafted features. However, the computation of convolutional features for each candidate window separately is computational expensive [Gir15].

In recent years, deep learning based detection frameworks like Faster R-CNN [RHGS15], which achieves top performing results on common detection benchmark datasets, have been proposed to reduce the computational effort. Therefore, a convolutional feature map is computed for the entire image at once and shared for all candidate windows [Gir15, RHGS15]. However, such detection frameworks are developed and optimized for common detection benchmark datasets that exhibit considerably differing characteristics compared to aerial images, e.g. size of objects to detect.

In the context of this report, we demonstrate the applicability of Faster R-CNN for vehicle detection in aerial images. Therefore, several adaptions are performed to account for the characteristics of the aerial images and the impact on the detection performance is evaluated. The DLR 3K Munich Vehicle Aerial Image Dataset [LM15] that comprises objects in the range of 15×30 pixels is used for all experiments.

The main improvement is achieved by adapting the resolution of the output of the last convolutional layer, which is used as feature map to localize and classify objects. The resolution of the standard feature map is only 1/16 of the input image and consequently insufficient for object sizes between 15 and 30 pixels. To provide a sufficient feature map resolution, the output of shallower convolutional layers is used as feature map. However, these features offer less semantic and contextual information compared to features of deeper layers. This results in more false alarms due to objects with similar shapes as vehicles. To account for that, we further propose a deconvolutional module that up-samples features of deeper layers and combines these features with features of shallower layers.

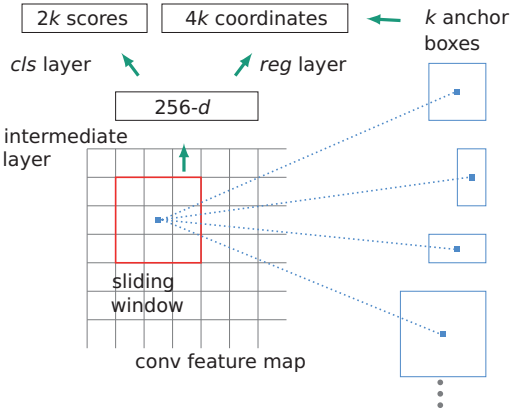


Figure 2.1: Schematic illustration of the Region Proposals Network (RPN) used to generate a set of candidate regions that are likely to contain an object.

2 Faster R-CNN

In the following, the functional principle of the Faster R-CNN detection framework as proposed by Ren et al. [RHGS15] is introduced. Faster R-CNN is comprised of two modules: an initial deep learning based object proposals method called Region Proposals Network (RPN) and the subsequent Fast R-CNN module [Gir15] used to classify the generated proposals. Both the RPN and the Fast R-CNN module share the convolutional layers to reduce the computational effort.

Figure 2.1 shows schematically the RPN. The RPN uses the output of the last convolutional layers as feature map. Then, a small network is shifted over the feature map to generate a set of candidate regions. The small network comprises a 3×3 convolutional layer followed by a classification layer (*cls* layer) and a bounding box regression layer (*reg* layer). The classification layer outputs a confidence score at each position, which is used to rank the proposals. The bounding box regression layer is used to compute the corresponding coordinates. For this, a set of fixed scaled anchor boxes k are used as bounding box reference.

The top 300 region proposals (highest confidence score) are forwarded to the Fast R-CNN module. The Fast R-CNN module classifies each region proposal into various object classes or background. Therefore, each region proposal is

projected onto the feature map. Then, the corresponding features are extracted by the so called Region of Interest (RoI) pooling layer to generate a vector of fixed length as required for the subsequent fully connected layers. After a sequence of fully connected layers, a classification layer and a bounding box regression layer are used for classification and to refine the coordinates of the corresponding candidate region, respectively.

3 Adaption to Aerial Images

The detection performance is mainly affected by adapting the resolution of the feature map used to compute proposals and for classification and by adapting the parameters of the RPN.

The original Faster R-CNN utilizes VGG-16 [SZ14] as base architecture. The VGG-16 comprises 13 convolutional layers with a kernel size of 3×3 followed by 3 fully-connected layers. To reduce the amount of parameters and to make the network invariant to small translations of the input, max-pooling layers are inserted after the 2nd (conv1_2), 4th (conv2_2), 7th (conv3_3), 10th (conv4_3), and 13th (conv5_3) convolutional layer. In case of Faster R-CNN, the output of the last convolutional layer is used as feature map. As illustrated in Figure 3.1, the dimensions of the feature map are only 1/16 of the dimensions of the input image. Thus, the feature map resolution is insufficient to accurately localize objects in the range of 15 to 30 pixels or even smaller. To account for that, we replace the initially used VGG-16 architecture by a network architecture optimized for handling small instances. The network is inspired by the network proposed in [HWB16] and comprises 4 convolutional layers followed by 3 fully connected layers. Max-pooling layers are inserted after the 1st, 2nd, and 4th convolutional layer. We performed optimization of all relevant network parameters including number of layers, number of filters per layer, kernel size and dropout. Analogous to the original Faster R-CNN, the output of the last convolutional layer is used as feature map. As depicted in Figure 3.2, the dimensions of the feature map are 1/4 of the dimensions of the input image. Thus, a finer localization of small objects is feasible due to the higher resolution of the feature map.

In addition to increasing the feature map resolution, adapting the parameters of the RPN mainly affects the detection performance. The benchmark datasets used

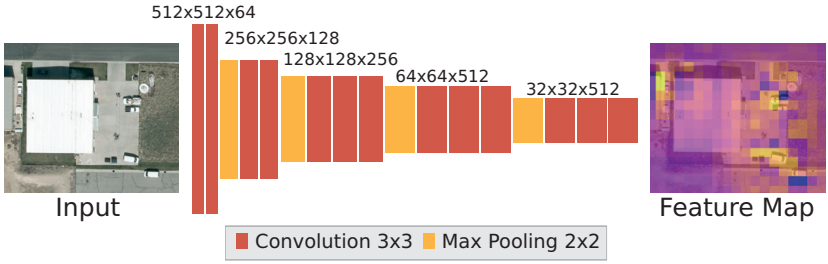


Figure 3.1: Schematic illustration of the convolutional part of VGG-16 and the resulting feature map.

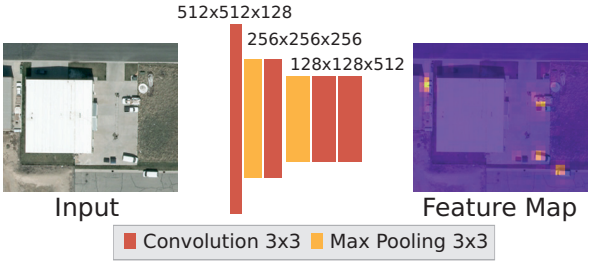


Figure 3.2: Schematic illustration of the convolutional part of the proposed network architecture optimized for handling small objects and the resulting feature map.

for developing Faster R-CNN contain objects that are generally in the range between 50 and 200 pixels. Thus, the parameters of the RPN are adjusted for these object dimensions. First, we reduce the minimal height and width of considered proposals (RPN_MI_SIZE) from 16 to 4, in order to account for the smaller object sizes in case of aerial images. Initially, the top 300 region proposals are considered for classification. This is enough to localize objects in the benchmark datasets, which generally contain only one or a few objects per image. Multiple proposals are typically located around the same object. Aerial images can contain clearly more objects per image and furthermore can contain more potentially disturbing objects, e.g. trailers or solar cells on buildings. Therefore, we set the number of proposals considered for classification to 2,000. As

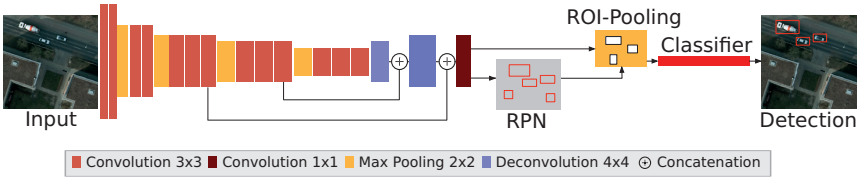


Figure 4.1: Schematic illustration of the Faster R-CNN extended by the deconvolutional module (DFRCNN).

described in Section 2, anchor boxes are used as reference for bounding box regression. The initially used anchor scales are chosen to account for the size of objects in the benchmark datasets. The ANCHOR_BASE_SIZE is set to 16 and the ANCHOR_SCALE factors are set to 8, 16 and 32, which results in anchor boxes with dimensions in the range between 128 and 512 pixels. We set the ANCHOR_BASE_SIZE to 2 while the ANCHOR_SCALE factors are kept unchanged.

4 Deconvolutional Module

To achieve a higher feature map resolution that is sufficient to localize small objects as in case of aerial images, small networks as described in Section 3 or shallow layers of standard architectures like VGG-16 are applicable. However, high-resolution feature maps offer less semantic and contextual information compared to features of deeper layers. The less semantic and contextual information make the detection framework more prone to false alarms due to objects with shapes similar to vehicles.

In order to achieve a high-resolution feature map and semantic and contextual informative features, we extend the Faster R-CNN by a deconvolutional module. The deconvolutional module up-samples low-dimensional feature maps of deep layers and combines the up-sampled features with the features of shallow layers while the feature map resolution is kept sufficiently high to localize small objects. The network architecture of the Faster R-CNN extended by the deconvolutional

module (DFRCNN) is schematically illustrated in Figure 4.1. We use VGG-16 as base network architecture. First, the features of conv5_3 are up-sampled by a factor of 2 and then concatenated with the features of conv4_3. Then, the combined features are up-sampled by a factor of 2 and then concatenated with the features of conv3_3. Thus, the features of conv4_3 and conv5_3 are up-sampled by a factor of 2 and 4, respectively. We use deconvolutional layers with a kernel size of 4×4 and a stride of 2 to up-sample the features. The combined features of conv3_3, conv4_4 and conv5_3 are used as feature map. The feature map dimensions are 1/4 of the dimensions of the input image. To adapt the number of output channels of the feature map required as input for the fully connected layers, we insert an additional convolutional layer with kernel size 1×1.

5 Evaluation

In the following section, we evaluate the impact of the adaptions described in Section 3 and of the deconvolutional module proposed in Section 4. We use Average Precision (AP) computed as defined in [EVGW+10], precision and recall as evaluation metrics. Ground truth (GT) objects are considered as recalled, if the Pascal-overlap criterion [EVGW+10] is satisfied. For all experiments, we use the publicly available DLR 3K Munich Vehicle Aerial Image Dataset. The dataset comprises 20 aerial images with a resolution of 5616×3744 pixels and a ground sampling distance (GSD) of approximately 13 cm. Due to the limited memory capacity of the used GPUs, each image is divided into tiles of 936×624 pixels. Image sections are exemplarily depicted in Figure 5.1. We further align the provided GT annotations at image edges as required for the Faster R-CNN detection framework.

5.1 Adaption to Aerial Images

The impact of increasing the feature map resolution on the detection performance is shown in Figure 5.2. The blue line corresponds to the precision-recall curve for an IoU threshold value of 0.5 used to accept GT objects as recalled (PASCAL-criterion). In case of using the VGG-16 architecture (feature map 1/16), both precision and recall are considerably worse compared to using the optimized network architecture (feature map 1/4). Precision values close to 1 and recall values

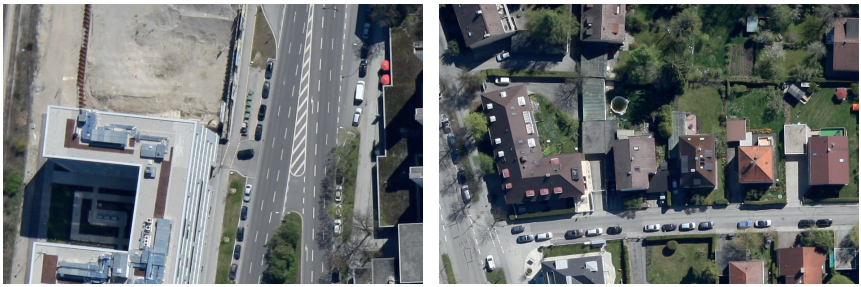


Figure 5.1: Image sections of the DLR 3K Munich Vehicle Aerial Image Dataset [LM15]

above 0.95 are achieved for the optimized network architecture. Reason for the improved performance is the higher feature map resolution as the detections are better localized around the GT elements. The better localization of the detections is illustrated in Figure 5.2 by plotting precision-recall curves for various IoU threshold values used to accept GT objects as recalled. For a resolution of 1/4 of the input image, the performance is only slightly decreasing with increasing IoU thresholds up to 0.5, which indicates a good localization of the detections. In contrast, the performance for lower resolutions decreases stronger with increasing IoU threshold values. The worse localization results in worse classification into object or background though the features comprise more semantic and contextual information. To highlight the difference in localization quality, qualitative detection examples are given for both feature map resolutions (see Figure 5.3 and Figure 5.4, respectively). For a resolution of 1/16 of the input image, the bounding box positions of the detections (red boxes) clearly differ from the GT annotations (green boxes). Furthermore, multiple detections are often generated due to the poor localization. In contrast, the detections for a feature map resolution of 1/4 of the input image overlap very well with the GT annotations.

The impact of adapting the RPN is illustrated in Figure 5.5 and Figure 5.6. Figure 5.5 depicts the proposals' quality for various anchor box sizes. Therefore, we plot the recall achieved for the proposals with respect to the IoU threshold value used to accept the GT objects as recalled. The mean anchor box dimensions are given in the legend. Reducing the anchor box sizes clearly improves the proposals' quality. For anchor box dimensions in the range between 14 and 28 pixels, which is roughly the size of present objects, the best recall values are

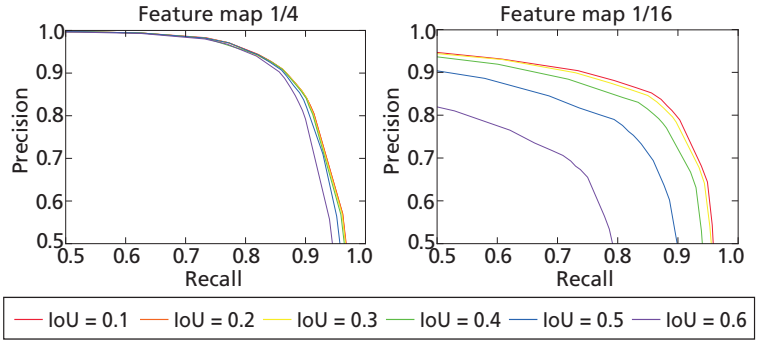


Figure 5.2: Precision-recall curves for various IoU threshold values used to accept GT objects as recalled. Higher feature map resolutions result in better localization quality as the performance decreases clearly less with increasing threshold values.

Figure 5.3: Qualitative detections (red boxes) and corresponding GT annotations (green boxes) for a feature map resolution of 1/16 of the input image. The detections show a relatively poor overlap with the GT annotations and multiple detections are often generated for one GT object.

achieved. The relation between proposals' quality and detection performance is shown in Figure 5.6. Therefore, we plot AP with respect to Average Best Overlap

Figure 5.4: Qualitative detections (red boxes) and corresponding GT annotations (green boxes) for a feature map resolution of 1/4 of the input image. The detections overlap very well with the GT annotations.

(ABO), which is an evaluation metric for the localization quality. ABO is calculated by averaging the best overlap between each GT annotation $g_i \in G$ and the corresponding set of object proposals L:

$$ABO = \frac{1}{|G|} \sum_{g_i \in G} \max_{l_j \in L} IoU(g_i, l_j) \,.$$

The best ABO is achieved for anchor box sizes in the range of present objects. The best AP is achieved for anchor boxes in the range of present objects as well. Thus, we assume that better proposals result in better detection performance.

To sum up the impact of the adaptions, the detection performance for both adaptions and the original Faster R-CNN is given in Figure 5.7. The performance of the original Faster R-CNN is poor. Both precision and recall are clearly less than 1. Applying the adapted RPN results in clearly improved precision and recall (VGG-16 adapted). However, the detection performance is still poor. Replacing the VGG-16 architecture with the optimized network architecture and consequently increasing the feature map resolution results in a significantly improved detection performance. It is to mention, that both adaptions are necessary to achieve the best detection results.

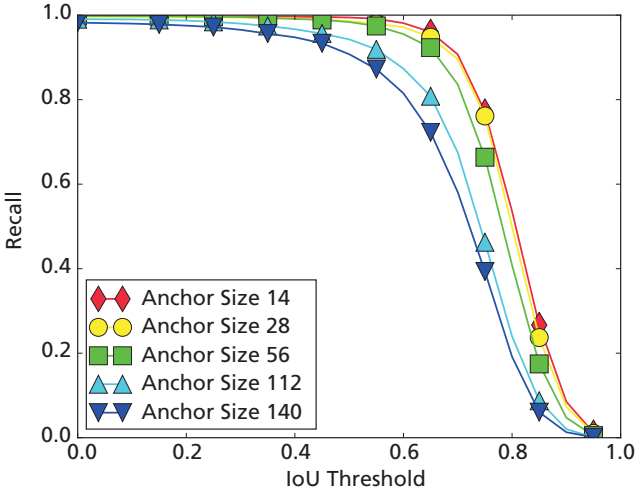


Figure 5.5: Recall-IoU curves for various anchor box sizes used as bounding box reference. The mean anchor box dimensions are given in the legend.

5.2 Deconvolutional Module

The impact of our proposed deconvolutional module on the detection performance is given in Table 5.1. We compare the detection results of our proposed Deconvolutional Faster R-CNN (DFRCNN) to baselines on the DLR 3K dataset for various GSDs. For this, we re-scaled the input images for training and testing by factors 1, 0.75, and 0.5. As baseline, we consider Faster R-CNN with different convolutional layers of VGG-16 used as feature map. For each GSD, the anchor box sizes are adapted for all Faster R-CNNs to the size of present objects. As discussed above increasing the feature map resolution from 1/16 of the input image (VGG-16 – conv5_3) to 1/4 of the input image (VGG-16 – conv3_3) clearly improves the detection performance especially for tiny objects as for a GSD of 26 cm. The performance is improved though the used features are less semantically and contextually informative. To account for the smaller receptive fields and less semantic information, we use our DFRCNN which combines features of conv3_3, conv4_3, and conv5_3 as described in Section 4. The performance is improved for all GSD especially for a GSD of 26 cm and consequently smaller

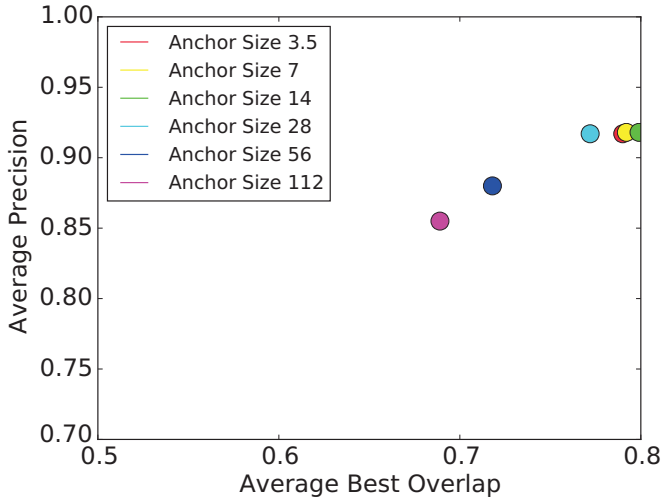


Figure 5.6: Average Precision (AP) w.r.t. Average Best Overlap (ABO) for various anchor box sizes used as bounding box reference. Applying region proposals with better ABO results in higher AP.

objects. In case of GSD 26, the number of false positive detections is reduced by a factor of 33.4% compared to VGG-16 — conv3_3, while the number of false negative detections remains almost unchanged.

To illustrate the impact of adding more semantic information, qualitative detection examples for Faster R-CNN using conv3_3 as feature map (left column) and

Table 5.1: Average Precision of our proposed DFRCNN compared to baselines on the DLR 3K dataset for various GSDs (in cm).

Method	GSD 13	GSD 19.5	GSD 26
VGG-16 — conv5_3	0.770	0.558	0.207
VGG-16 — conv4_3	0.967	0.896	0.601
VGG-16 — conv3_3	0.979	0.944	0.836
DFRCNN	**0.980**	**0.957**	**0.864**

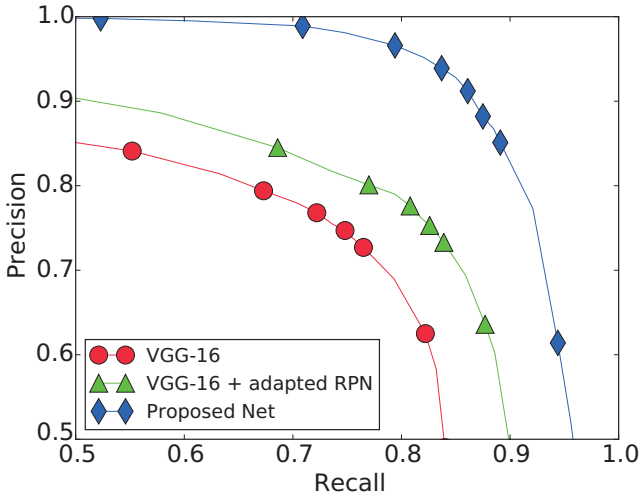


Figure 5.7: Precision-recall curves of the original Faster R-CNN and for both adaptions.

our proposed DFRCNN (right column) are given in Figure 5.8. Therefore, we use a classification threshold value of 0.5. For Faster R-CNN using conv3_3, several false positive detections are caused by objects with shapes similar to vehicles such as solar cells or chimneys on buildings. Integrating more semantic information clearly reduces the number of false positive detections caused by such objects.

6 Summary

In this report, the applicability of Faster R-CNN for vehicle detection in aerial images was demonstrated. Therefore, we have systematically evaluated the impact of adapting relevant parameters of Faster R-CNN to the characteristics of aerial images. The most improvement in detection performance was achieved by adapting the size of the anchor boxes used for bounding box regression and by increasing the feature map resolution as the initial resolution is insufficient to localize small objects. To achieve high feature map resolutions that are sufficient to localize small objects, small networks or shallow layers of standard architectures

Figure 5.8: Qualitative detections (red boxes) and corresponding GT (green boxes) for Faster R-CNN using *conv3_3* (left column) and our proposed DFRCNN (right column) on DLR 3K indicate that false alarms due to objects with shapes similar to vehicles are reduced by integrating more semantic information.

like VGG-16 are applicable, which offer less semantic and contextual information compared to features of deeper layers. In order to overcome this drawback, we extended the original Faster R-CNN by a deconvolutional module. Therefore, features of deeper layers are up-sampled and combined with features of shallower layers. The detection performance is improved by integrating features with more semantic information especially for tiny objects as the number of false positive detections due to objects with shapes similar to vehicles is reduced.

Bibliography

[CH16] Gong Cheng and Junwei Han. A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117:11–28, 2016.

[CXLP14] Xueyun Chen, Shiming Xiang, Cheng-Lin Liu, and Chun-Hong Pan. Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE GRSL*, 11(10):1797–1801, 2014.

[EVGW+10] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[Gir15] Ross Girshick. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1440–1448, 2015.

[HWB16] Christian Herrmann, Dieter Willersinn, and Jürgen Beyerer. Low-resolution convolutional neural networks for video face recognition. In *Advanced Video and Signal Based Surveillance*. IEEE, 2016.

[KPF16] Georgy V Konoplich, Evgeniy O Putin, and Andrey A Filchenkov. Application of deep learning to the problem of vehicle detection in UAV images. In *Soft Computing and Measurements (SCM), 2016 XIX IEEE International Conference on*, pages 4–6. IEEE, 2016.

[LM15] K. Liu and G. Mattyus. Fast multiclass vehicle detection on aerial images. *GRSL, IEEE*, PP(99):1–5, 2015.

[MM14] Thomas Moranduzzo and Farid Melgani. Detecting cars in UAV images with a catalog-based approach. *IEEE Transactions on Geoscience and Remote Sensing*, 52(10):6356–6367, 2014.

[RHGS15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[SZ14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Karlsruher Schriftenreihe zur Anthropomatik (ISSN 1863-6489)

Herausgeber: Prof. Dr.-Ing. habil. Jürgen Beyerer

Band 1 Jürgen Geisler
Leistung des Menschen am Bildschirmarbeitsplatz. 2006
ISBN 3-86644-070-7

Band 2 Elisabeth Peinsipp-Byma
Leistungserhöhung durch Assistenz in interaktiven Systemen zur Szenenanalyse. 2007
ISBN 978-3-86644-149-1

Band 3 Jürgen Geisler, Jürgen Beyerer (Hrsg.)
Mensch-Maschine-Systeme. 2010
ISBN 978-3-86644-457-7

Band 4 Jürgen Beyerer, Marco Huber (Hrsg.)
Proceedings of the 2009 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory. 2010
ISBN 978-3-86644-469-0

Band 5 Thomas Usländer
Service-oriented design of environmental information systems. 2010
ISBN 978-3-86644-499-7

Band 6 Giulio Milighetti
Multisensorielle diskret-kontinuierliche Überwachung und Regelung humanoider Roboter. 2010
ISBN 978-3-86644-568-0

Band 7 Jürgen Beyerer, Marco Huber (Hrsg.)
Proceedings of the 2010 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory. 2011
ISBN 978-3-86644-609-0

Band 8 Eduardo Monari
Dynamische Sensorselektion zur auftragsorientierten Objektverfolgung in Kameranetzwerken. 2011
ISBN 978-3-86644-729-5

Die Bände sind unter www.ksp.kit.edu als PDF frei verfügbar oder als Druckausgabe bestellbar.

Band 9 Thomas Bader
Multimodale Interaktion in Multi-Display-Umgebungen. 2011
ISBN 3-86644-760-8

Band 10 Christian Frese
Planung kooperativer Fahrmanöver für kognitive Automobile. 2012
ISBN 978-3-86644-798-1

Band 11 Jürgen Beyerer, Alexey Pak (Hrsg.)
Proceedings of the 2011 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory. 2012
ISBN 978-3-86644-855-1

Band 12 Miriam Schleipen
Adaptivität und Interoperabilität von Manufacturing Execution Systemen (MES). 2013
ISBN 978-3-86644-955-8

Band 13 Jürgen Beyerer, Alexey Pak (Hrsg.)
Proceedings of the 2012 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory. 2013
ISBN 978-3-86644-988-6

Band 14 Hauke-Hendrik Vagts
Privatheit und Datenschutz in der intelligenten Überwachung: Ein datenschutzgewährendes System, entworfen nach dem „Privacy by Design" Prinzip. 2013
ISBN 978-3-7315-0041-4

Band 15 Christian Kühnert
Data-driven Methods for Fault Localization in Process Technology. 2013
ISBN 978-3-7315-0098-8

Band 16 Alexander Bauer
Probabilistische Szenenmodelle für die Luftbildauswertung. 2014
ISBN 978-3-7315-0167-1

Band 17 Jürgen Beyerer, Alexey Pak (Hrsg.)
Proceedings of the 2013 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory. 2014
ISBN 978-3-7315-0212-8

Die Bände sind unter www.ksp.kit.edu als PDF frei verfügbar oder als Druckausgabe bestellbar.

Band 18 Michael Teutsch
Moving Object Detection and Segmentation for Remote Aerial Video Surveillance. 2015
ISBN 978-3-7315-0320-0

Band 19 Marco Huber
Nonlinear Gaussian Filtering: Theory, Algorithms, and Applications. 2015
ISBN 978-3-7315-0338-5

Band 20 Jürgen Beyerer, Alexey Pak (Hrsg.)
Proceedings of the 2014 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory. 2014
ISBN 978-3-7315-0401-6

Band 21 Todor Dimitrov
Permanente Optimierung dynamischer Probleme der Fertigungssteuerung unter Einbeziehung von Benutzerinteraktionen. 2015
ISBN 978-3-7315-0426-9

Band 22 Benjamin Kühn
Interessengetriebene audiovisuelle Szenenexploration. 2016
ISBN 978-3-7315-0457-3

Band 23 Yvonne Fischer
Wissensbasierte probabilistische Modellierung für die Situationsanalyse am Beispiel der maritimen Überwachung. 2016
ISBN 978-3-7315-0460-3

Band 24 Jürgen Beyerer, Alexey Pak (Hrsg.)
Proceedings of the 2015 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory. 2016
ISBN 978-3-7315-0519-8

Band 25 Pascal Birnstill
Privacy-Respecting Smart Video Surveillance Based on Usage Control Enforcement. 2016
ISBN 978-3-7315-0538-9

Band 26 Philipp Woock
Umgebungskartenschätzung aus Sidescan-Sonardaten für ein autonomes Unterwasserfahrzeug. 2016
ISBN 978-3-7315-0541-9

Band 27 Janko Petereit
Adaptive State × Time Lattices: A Contribution to Mobile Robot Motion Planning in Unstructured Dynamic Environments. 2017
ISBN 978-3-7315-0580-8

Band 28 Erik Ludwig Krempel
Steigerung der Akzeptanz von intelligenter Videoüberwachung in öffentlichen Räumen. 2017
ISBN 978-3-7315-0598-3

Band 29 Jürgen Moßgraber
Ein Rahmenwerk für die Architektur von Frühwarnsystemen. 2017
ISBN 978-3-7315-0638-6

Band 30 Andrey Belkin
World Modeling for Intelligent Autonomous Systems. 2017
ISBN 978-3-7315-0641-6

Band 31 Chettapong Janya-Anurak
Framework for Analysis and Identification of Nonlinear Distributed Parameter Systems using Bayesian Uncertainty Quantification based on Generalized Polynomial Chaos. 2017
ISBN 978-3-7315-0642-3

Band 32 David Münch
Begriffliche Situationsanalyse aus Videodaten bei unvollständiger und fehlerhafter Information. 2017
ISBN 978-3-7315-0644-7

Band 33 Jürgen Beyerer, Alexey Pak (Eds.)
Proceedings of the 2016 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory. 2017
ISBN 978-3-7315-0678-2

Band 34 Jürgen Beyerer, Alexey Pak and Miro Taphanel (Eds.)
Proceedings of the 2017 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory. 2018
ISBN 978-3-7315-0779-6

Lehrstuhl für Interaktive Echtzeitsysteme
Karlsruher Institut für Technologie

Fraunhofer-Institut für Optronik, Systemtechnik
und Bildauswertung IOSB Karlsruhe

In 2017, the annual joint workshop of the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB) and the Vision and Fusion Laboratory (IES) of the Institute for Anthropomatics, Karlsruhe Institute of Technology (KIT) has again been hosted by the town of Triberg-Nussbach in Germany. For a week from July, 30 to August, 5 the doctoral students of both institutions presented extensive reports on the status of their research and discussed topics ranging from computer vision and optical metrology to network security and machine learning. The results and ideas presented at the workshop are collected in this book in the form of detailed technical reports. This volume provides a comprehensive and up-to-date overview of the research program of the IES Laboratory and the Fraunhofer IOSB.

ISSN 1863-6489
ISBN 978-3-7315-0779-6

Gedruckt auf FSC-zertifiziertem Papier