

GFaI Tagungsband 2023



AI4EA Workshop (Berlin Workshop on Artificial Intelligence for Engineering Applications)

3D-NordOst (3D in Science & Applications)



Gesellschaft zur Förderung
angewandter Informatik e.V.

INHALTSVERZEICHNIS

GFal-Tagungsband 2023

VORWORT

2 Vorwort der Workshopleitung

PUBLIKATIONEN AI4EA

- 6 Vibration Analysis of Sawing Processes aided by AI
- 13 AI-based Wear Analysis of Sawing Processes using Machine Data
- 20 TIADE: A Toolbox for Identifying and Analyzing Anomaly Detection Errors
- 26 Substitution of FEM simulations aided by AI
- 33 Classification of stone samples for concrete production using hyperspectral imaging and deep learning

PUBLIKATIONEN 3D-NordOst / 3DiSA

- 40 ELV.xD - a modular framework for instance level recognition within spare part logistics.
- 46 Segment Anything Model für Zero-Shot 3D Bounding Box Prediction
- 55 Informationsaustausch zwischen einer MixedReality-Anwendung für Stadtplanungsprozesse und BIM-Systemen mit dem Standard BCF
- 61 Comparison of Digital Model with Model in Production for Quality Assurance Purposes
- 69 Vergleich von optischen 3D-Sensoren zur Feinreferenzierung eines Klebprozesses für schiffbauliche Komponenten
- 76 Virtuelles Fitting zur Unterstützung bei der Entwicklung von neuartigen Kanülen zur mechanischen Kreislaufunterstützung von Patienten mit univentrikulären Herzen
- 79 KI-gestützte Umsetzung für den digitalen Prozess von 3D-gedruckten Orthesen
- 81 Evaluation of Depth Completion Using Solid-State LiDAR Dataset
- 88 Entwicklung eines Verfahrens zur auflösungsangepassten Abschätzung der Schichtdicke in einer modellbasierten Auftragssimulation durch adaptive Granularisierung
- 93 Photogrammetrische Bestimmung von Kopfbewegungen mittels Facial Landmarks – lessons learned
- 100 Identifikation klinisch interpretierbarer Feature-Vektoren für die Analyse von Torso Surface-Topographie Daten

VORWORT



AI4EA

Künstliche Intelligenz prägt Wissenschaft, Industrie und Technik, wenn es um modernste Technologien geht. Von der Produktentwicklung über die Produktion bis hin zur Wartung und Zustandsüberwachung schafft Künstliche Intelligenz Lösungen, die noch vor wenigen Jahren unerreichbar schienen.

Der Berliner Workshops zu Künstliche Intelligenz für Ingenieur-Anwendungen (AI4EA) setzt bei diesen Themen an und stellt KI-basierte Lösungen für industrielle Probleme vor. Er fand dieses Jahr bereits zum zweiten Mal statt und wir freuen uns über das wachsende Interesse an dieser Veranstaltung als Plattform für den Wissenstransfer. Denn der Workshop fokussiert sich explizit auf den Wissensaustausch von Forschung und Industrie. AI4EA bietet sowohl Wissenschaftlern, als auch Industriefachleuten die Möglichkeit, ihre Ideen, Ergebnisse und Probleme zu diskutieren und zu veröffentlichen. In diesem Jahr konzentrierte sich AI4EA auf die Anwendungsbereiche Qualitätssicherung in der Produktion, Predictive Maintenance, Prozesssteuerung/-optimierung und Produktentwicklung.

Wir haben insgesamt 13 Einreichungen erhalten, von denen alle für die Präsentation akzeptiert wurden. Da die Autoren die auch Möglichkeit hatten alternativ nur ein Poster zu präsentieren und gleichzeitig nicht alle zur Veröffentlichung eingereichten Beiträge auch zugelassen wurden, ist die Anzahl der Veröffentlichungen in diesem Band mit fünf Beiträgen geringer. Jeder zur Veröffentlichung eingereichte Beitrag durchlief eine Begutachtungsphase mit mindestens zwei, meist jedoch drei oder vier anonymen Reviews. Wir danken dem Programmkomitee für die engagierte Begutachtung der Einreichungen. Insbesondere danken wir Frau Beuster vom GRW-Netzwerk NET4AI und Herrn Hohnhäuser vom ZIM-Innovationsnetzwerk AI4Tech als Mitveranstalter des Workshops für die Moderation und Unterstützung. Abschließend danken wir den Autoren für ihre Mühe beim Verfassen ihrer Beiträge und für die hervorragenden Präsentationen.

November 2023

Benny Botsch
Dr. Daniel Herfert
Mario Koddenbrock
Miriam Schneider

Programm Komitee

Hannes Diethelm	GFal e.V., Germany
Wolfgang A. Halang	FernUniversität Hagen, Germany
Alfred Iwainky	GFal e.V., Germany
Benjamin Noack	Karlsruher Institut für Technologie, Germany
Christof Puhle	GFal e.V., Germany
Jörg Reiff-Stephan	Technische Hochschule Wildau, Germany
Erik Rodner	Hochschule für Technik und Wirtschaft Berlin, Germany
Holger Schlingloff	Humboldt-Universität zu Berlin, Germany
Gerd Stanke	GFal e.V., Germany

VORWORT



3D-NordOst

In diesem Jahr konnten wir uns über das 25. Jubiläum des anwendungsbezogenen Workshops zur Erfassung, Modellierung, Verarbeitung und Auswertung von 3D-Daten freuen. Ein viertel Jahrhundert voller Geschichte und Tradition aber auch voller Innovationen, Forschungsergebnissen und technischen Neuerungen. Neben der Wahrung bewährter Traditionen wie der Praxishäufigkeit des Workshops, der Vielseitigkeit der Themengebiete und des unkomplizierten Einreichungs- und Teilnahmeprozesses, wurden dieses Jahr einige notwendige Neuerungen eingeführt. Erstmals konnte allen Autoren auf freiwilliger Basis ein Review-Prozess (Peer-Review Verfahren) angeboten werden, welcher die Qualität der Druckbeiträge sichert und weiter erhöht. Um die Verfügbarkeit der wissenschaftlichen Beiträge zu bewahren und weiter zu verbessern wird neben der bewährten gebundenen Druckausgabe erstmals ein Tagungsband als digitale Online Version auf der Konferenz-Webseite als PDF-Version veröffentlicht. Tradition und Geschichte sind wichtige Säulen, jedoch liegt uns auch der wissenschaftliche Nachwuchs am Herzen. Aus diesem Grund wurde in der 25. Ausgabe der 3D-NordOst erstmals ein Studenten-Pitch durchgeführt, bei dem Studenten und Nachwuchswissenschaftler ihre Forschungsthemen in einem Kurzformat (5 Minuten) vortragen konnten. Um alle Aktivitäten des Konferenz- und Verlagsmanagements weiter zu bündeln und zu professionalisieren, wurde die Dachmarke ‚3Disa, 3D in science & applications‘ gegründet. Dennoch bleibt die 3D-NordOst als traditionsreicher Namensgeber auch in Zukunft bestehen. Ein wesentlicher Trend der auch die 3D-Datenverarbeitung durchdringt ist natürlich das Thema künstliche Intelligenz (KI). Etwa fünfzig Prozent aller Beiträge in diesem Jahr beschäftigten sich direkt oder indirekt mit den Methoden der KI. Aus diesem Grund wird die 3D-NordOst im Jahr 2024 erstmals zusammen mit dem ebenfalls hauseigenen auf das Thema KI spezialisierten Workshop AI4EA ausgerichtet. Als Vorbereitung dazu und zur Erweiterung des Leserkreises wird der diesjährige Tagungsband bereits zusammen als gemeinsame erweiterte Ausgabe mit dem Workshop AI4EA veröffentlicht.

November 2023

Dr.-Ing. Stephan Rothstock

Programm Komitee

Felicitas Böhm	GFal e.V., Deutschland
Ben Hohnhäuser	GFal e.V., Deutschland
Daniel Krueger	GFal e.V., Deutschland
Frank Neumann	Hochschule für Technik und Wirtschaft Berlin, Deutschland
Stephan Rothstock	GFal e.V., Deutschland
Holger Schlingloff	Humboldt-Universität zu Berlin, Deutschland

Danksagung

Unser Dank gilt den Akteuren und Organisatoren beider Workshops vor allem dem Programm-Komitee und den zahlreichen Gutachtern sowie den helfenden Händen, die im Hintergrund zur Vorbereitung und zum Gelingen der Veranstaltung beigetragen haben.



AI4EA Workshop (Berlin Workshop on Artificial Intelligence for Engineering Applications)



Mitveranstalter:

AI4Tech (ZIM-Innovationsnetzwerk)

NET4AI (GRW-Netzwerk)

Berlin
16.11./17.11.2023
Gesellschaft zur Förderung
angewandter Informatik e. V.

Vibration Analysis of Sawing Processes aided by AI

Jens Kneifel¹, Christoph Böttinger¹, Robin Roj¹, Christian Pelshenke¹, Ralf TheiB¹ and Peter Dültgen¹

¹Forschungsgemeinschaft Werkzeuge und Werkstoffe e. V.,
Papenberger Str. 49,
42859 Remscheid, Germany
roj@fgw.de

Abstract. This paper presents the experimental application of an Industrial-Internet-of-Things prototype called NodeOne. It combines several tasks, such as the acquisition of high-frequency measurement data and the evaluation by machine learning methods in an artificial intelligence application. The installed measurement technology and open-source software generates equivalent measurement data as high-priced laboratory equipment, so that it can be used as a low-budget substitute. A workflow is designed that promotes experimental work and simplifies the effort required to implement artificial intelligence solutions. In addition to prove the functionality of the hardware and software, it was intended to test its applicability in an industrial context and for research tasks. For this purpose, the results of sawing experiments are presented that were conducted to collect measurement data, extract suitable features, and train artificial intelligence models. Techniques from vibration analysis were used for feature extraction, the results from artificial intelligence model creation were tested in a live application, and the resulting classification was evaluated. The NodeOne is designed for communication in a network and is compatible with typical systems. It is shown that the created system achieves sufficient accuracy in predictions and the designed process can be used for artificial intelligence rapid prototyping.

Keywords: Predictive Maintenance, Predictive Quality, Process Data, Sawing Processes, Vibration Analysis.

1 INTRODUCTION

In the context of digitalization, innovative technologies, which are summarized under the term „Industry 4.0“, are being used particularly in the area of production [1]. One area of Industry 4.0 is artificial intelligence (AI). While large companies have the personnel and financial resources to drive forward the digitalization of their processes in a targeted manner and to research problems, small and medium enterprises (SME) are often unable to recruit their own specialists or establish their own departments. Thus, SMEs are dependent on purchasing commercial standard solutions that do not meet the desired services or only to a limited extent.

The advantages of Industry 4.0 applications are obvious. For example, they enable a higher degree of automation, the reduction of waste, higher utilization of machines as well as shorter maintenance and downtimes. Overall, this results in an improvement in the cost-benefit ratio. As a result of these improvements, the higher investment costs should pay for themselves after only a short period of use.

One important possibility is the condition monitoring of machines. In recent years, progress has been made in the area of sensor technology, which allows automated acquisition of process data through interacting sensor networks. Especially SMEs depend on the use of innovative technologies to add value to their products and reduce labor costs. The monitoring of machines is an elementary base for further optimization. New approaches to equipping older machines as so-called retrofits are of particular importance, as these work with existing machines and therefore build on available know-how

and do not require new, expensive equipment.

The central aspect of this work is to collect machine and process data and evaluate it using AI. The provision of a suitable IT infrastructure and the gain of knowledge through automated data analysis are a prerequisite. The measurement hardware is to monitor different sensors and send the raw data to an arbitrary network. Here, the data is preprocessed so that AI-based analysis can be automated. The results are sent back so that they can be visualized on end devices. The aim is to provide SMEs with access, advice, and support on the topics of machine learning (ML) and AI so that they can implement these techniques in their companies. In this way, these companies will be able to prepare for the production standards of the future and optimize their processes. The goal thus results in condition monitoring of different processes. For this purpose, a sensor network is used, which is also suitable for older machines as retrofit. The main focus of the network is on a vibration analysis for structure-borne noise and generated vibrations. This allows workpiece quality to be predicted or maintenance intervals to be better planned. Ideally, downstream quality controls can be reduced. Consequently, existing resources are better utilized, and workflows are optimized, which leads to a higher competitiveness of the companies.

In order to lay the foundation, the current status for measurement data acquisition and sensor technology is considered in the following sections. Current research is concerned with different sensors that monitor the process. These are attached to the tool, workpiece, or machining center so that conclusions can be drawn about what is happening in the process [2-5].

2 MATERIALS AND METHODS

2.1 SENSOR TECHNOLOGY

Three different operating principles are presented below. Since one aim of this work is a low-cost entry for SMEs, the price was also considered. Table 1 provides an overview of the technical properties.

Piezoelectric acceleration sensors

Piezoceramics are characterized by the fact that they induce a voltage when they are mechanically deformed. If this voltage is measured, a conclusion can be drawn about the vibration. Piezoelectric acceleration sensors have a large dynamic range. Inexpensive sen-

sors already have a maximum frequency range of up to 10 kHz with a maximum acceleration of 50 g. They offer a non-linearity of the acceleration range of 1 %. They may be subjected to an acceleration of 50 g.

Piezoresistive

Piezoresistive ceramics change resistance when mechanical forces are applied to them. These changes can be used to infer vibrations. They are suitable for low and medium frequencies up to 4 kHz. Due to their maximum acceleration of up to 5,000 g, piezoresistive sensors are ideally suited for shock measurements. They are also characterized by a high sensitivity of 0.15 mV/g and a non-linearity of 1 %.

Capacitive

Capacitive sensors can be manufactured at very low cost. The measuring range reaches up to 7 kHz. They can be accelerated up to 400 g and have a sensitivity of 10 mV/g. The non-linearity is specified with 1 %.

Principle	piezoelectric	piezo-resistive	capacitive
Sensitivity	100 mV/g	0.15 mV/g	10 mV/g
Range	0.2 Hz-20 kHz	0-2 kHz	0-5 kHz
Linearity	1 %	1 %	1 %
Acceleration	50 g	6,000 g	400 g

Table 1. Comparison of sensor principles

Due to their properties, piezoelectric acceleration sensors are best suited for use in this work. Especially the large measuring range of inexpensive measuring sensors is decisive here. The maximum acceleration of 50 g is sufficient. For the implementation, the following criteria are set up, which are summarized under the keyword integrability.

Different sensors from the company PCB Piezotronics are compared directly with each other since this company manufactures the sensors directly [6]. They offer the best available integration options with different attachments for their sensors (screws metric/imperial, magnet), as well as the best possible protection against external influences, since they comply with the military standard. Therefore, different sensors from this company were preselected and compared in Table 2.

Piezoelectric sensors are especially in connection with the IEPE standard outstanding for the requirements in this work. By using the IEPE standard, through which the signals are transmitted with low impedance, it is possible to place the sensor and the data logger at a greater local

distance from each other. This, combined with a very wide measurable frequency range, gives a technology advantage for these sensors over capacitive and piezoresistive sensors.

Table 2. Sensor properties

Sensor	PCB-M621C40	PCB-M601A02	PCB-603M170	ICP 356A02
<i>Main data</i>				
Sensitivity	10 mV/g	500 mV/g	100 mV/g	10 mV/g
Range	500 g	10 g	80 g	500 g
Frequency range	1.6-30,000 Hz	0.47-4,000 Hz	0.4-14,000 Hz	1-5,000 Hz
Broadband resolution	1.176 $\mu\text{m/s}^2$	343 $\mu\text{m/s}^2$	5,395 $\mu\text{m/s}^2$	5 $\mu\text{m/s}^2$
Linearity error	1 %	1 %	1 %	1 %
<i>Operating conditions</i>				
Overload limit	10,000 g	5,000 g	5,000 g	7,000 g
Temperature range	-54 - 121 °C	-54 - 121 °C	-54 - 121 °C	-54 - 121 °C
<i>Technical properties</i>				
Settling time	< 3.0 s	< 10 s	< 2.0 s	< 5 s
Discharge time constant	> 0.1 s	> 1 s	> 0.4 s	< 2.0 s

2.2 MEASURING HARDWARE

In the case of the sensors, the range from which they can be selected is very large. The ICP 356A02 fulfills the most requirements and was hence selected for the application. One focus is to offer additional hardware options to reduce the initial investment and thus the entry barrier for SMEs. Therefore, a measurement hardware for data acquisition has been developed. It is called NodeOne in the following [7]. It uses a Raspberry-Pi4 in combination with an attached measuring hardware in a MaticBox4. This has the advantage of a rapid development, a very small form factor, low power consumption and low costs. As a measuring card an MCC-172 is installed. This has two connections specifically for IEPE (piezo) sensors. The system is modular, so that several measuring cards can be used. This has the further advantage that the data logger is scalable and can therefore be adapted very well to the task.

Python was chosen as the programming language. It offers many possibilities, especially with

regard to ML and AI processes, as many libraries and frameworks are available as an open-source version. Consequently, no further license fees are necessary. In comparison with commercially available lab grade solutions, the use of NodeOne results in savings of up to 94 %. Particularly noteworthy is the fact that an open-source solution has been developed, which means that the ongoing costs for licenses are completely eliminated.

2.3 EXPERIMENTAL SETUP

An industrial metal band saw of the type „e-cut 400A“ from the company MEBA Metall-Bandsäge-maschinen GmbH was equipped with an IEPE acceleration sensor ICP 356A02 at the rigid clamping jaw (cf. Fig. 1) [8]. Cuts were made on workpieces with different geometries and measurement data was acquired. An AI model was trained to recognize the geometry based on the shrinkage signals.



Figure 1: MEBA-Bandsäge e-cut 400A mit Sensor ICP 356A02 der Firma PCB.

The first workpiece was a square steel tube with dimensions 40 by 40 mm and represents the class „MEBA_1_square_hollow“ (class1). The second workpiece was a round steel tube in solid material with the diameter 40 mm and represents the class „MEBA_2_round_full“ (class2). In each case, 10 mm wide cuts were made at a feed rate of 1 mm per second, so that one cut took exactly 40 s (cf. Fig. 2).

The ICP 356A02 has three axes, of which the Y axis was measured along the direction of movement of the saw band. The sensitivity of 9.87 mV/g was also taken from the calibration sheet and entered in the parameter file in the measuring box. In Fig. 1, the NodeOne can be seen on the bottom right, next to another Raspberry Pi, which serves as the edge server here. The white box is a router that provides its own network for the experimental setup. Also, the placement of the NodeOne in the control cabinet can be recognized, where an installation on a top-hat rail was tested. The two workpieces are depicted in Fig. 2, where the top shows them during the sawing process and the bottom in a cut-off state. Seven cuts were made per class and workpiece type, respectively. For the classification, signal courses with a length of 1 s at a measuring frequency of 5,000 Hz were examined, so that a total of 280 data points per class were generated from the seven cuts of 40 seconds each.

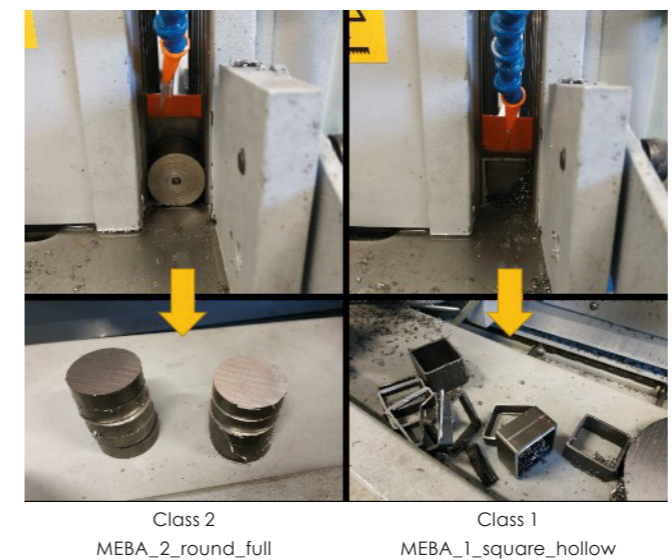


Figure 2: MEBA-Bandsäge e-cut 400A mit Sensor ICP 356A02 der Firma PCB.

Fig. 3 visualizes the process of data acquisition, processing, and evaluation. For that, the NodeOne and a Raspberry Pi were set up and connected to the network and the attached sensor. For this purpose, the setup was designed to allow live monitoring of the sawing test, live display of the neural network, and live image transmission to observe the sawing experiment.

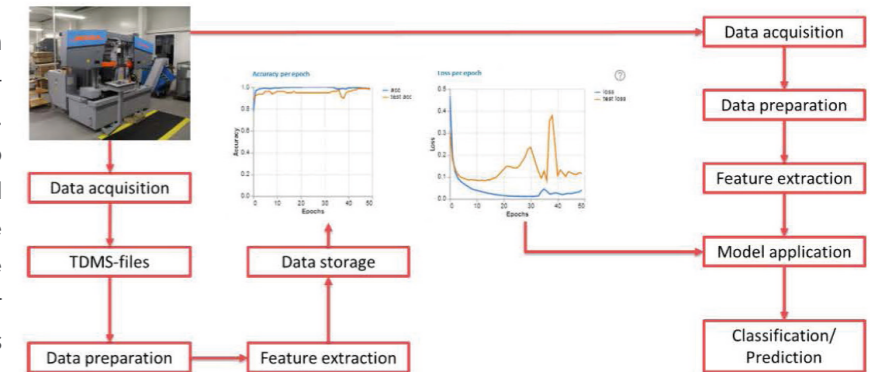


Figure 3: Flow chart of data acquisition and processing.

2.4 EXECUTION OF THE EXPERIMENT

A dataset of spectrograms was created for the AI model training. During feature extraction, it was found that the start and end points were set too tightly and a total of 9.64 % fewer data points than anticipated were created for the AI training. Ultimately, there were 272 data points for the class „MEBA_1_square_hollow“ (class1) and 234 data points for the class „MEBA_2_round_full“ (class2).

The measurement files were again converted into a data set with two classes using the „main_feature_batch“ function. The vibration signals were converted into spectrograms with a dimension of 129 frequencies times 22 timesteps. The setting was chosen because the measurement signal remains relatively constant over the time course of 1 s (no sporadic strong frequency changes) and therefore a higher resolution on the frequency axis was advantageous. For the creation of the spectrogram images the shading parameter „gouraud“ within the matplotlib library was used for better extrapolation while resizing as it interpolates between given datapoints instead of displaying the same values over multiple pixels. Fig. 4 shows the „gouraud“ data set with two classes in the AI transfer learning in Teachable Machine [9], which in itself relies on a pretrained network called MobileNet which is mainly used for image recognition [10].

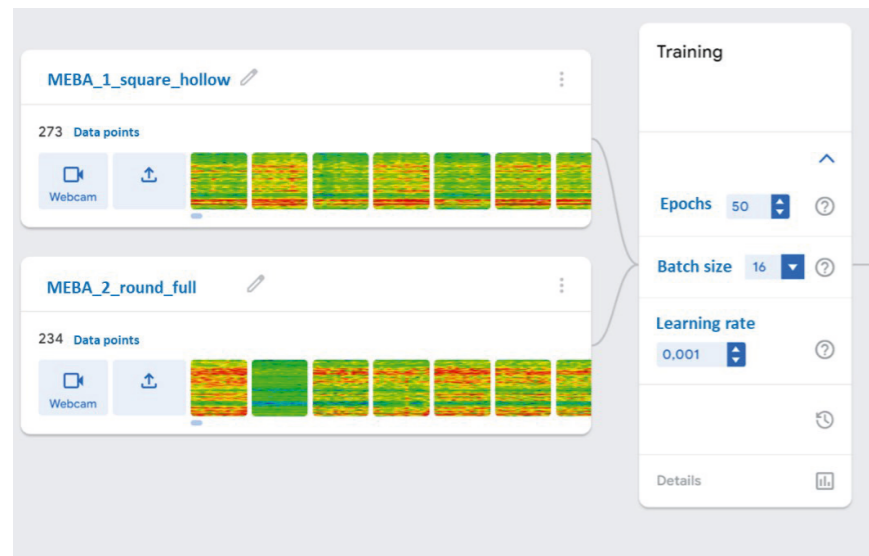


Figure 4: Overview of the „gouraud“-model.

The settings from the AI training were the same in training 1 and 2. In training 3, the settings have been changed. The AI transfer learning was performed in training 1 and 2 with the following parameters:

- Epochs: 50 (number of runs of the learning process)
- Batch size: 16 (number of training samples in a forward/backward pass)
- Learning rate: 0.001 (adaption speed to new input)

The AI transfer learning in training 3 was performed with the following parameters:

- Epochs: 50
- Batch size: 32
- Learning rate: 0.00105

Thus, a total of three training runs were performed for the data set and the results were compared. For the comparison, the output of Tensorboard was used, which shows statistics from the learning process of Tensorflow within Teachable Machine.

3 RESULTS

The metrics on the learning process depict that the input data was split for „MEBA_1_square_hollow“ (class1) with 232 data points for training and 41 for evaluation as well as for „MEBA_2_round_full“ (class2) with 198 data points for training and 36 for evaluation. Fig. 5 shows the statistics from model training 1 and 2. In training 1, the „loss“ values increased tremendously between epochs 20 and 40, while the accuracy „acc“ remains putatively high. This suggests that the model is still in the region of overfitting. In training 2, the results are already better, with a slight increase in the „loss“ values in epochs 40

to 50, although there is no decrease in the „loss“ values in the range of epochs 10 to 40. This suggests that the model training was provided with a too small data set. However, the final values for the accuracy per class and the confusion matrix show an acceptable result with accuracies between 98 % and 99.99 %.

Since the deflections in the course of the „loss“ values in training 1 and training 2 occurred unexpectedly late, slightly modified parameters for transfer learning are applied for training 3. The batch size was increased so that more data was available during the forward/backward passes in the training process.

In addition, the learning rate was minimally increased because the dataset is very small, and this causes the AI training to make adjustments to the network more quickly. This value had to be adjusted very carefully, because even with small changes an overfitting can occur quickly, which cannot be overcome by more epochs due to the small amount of data available.

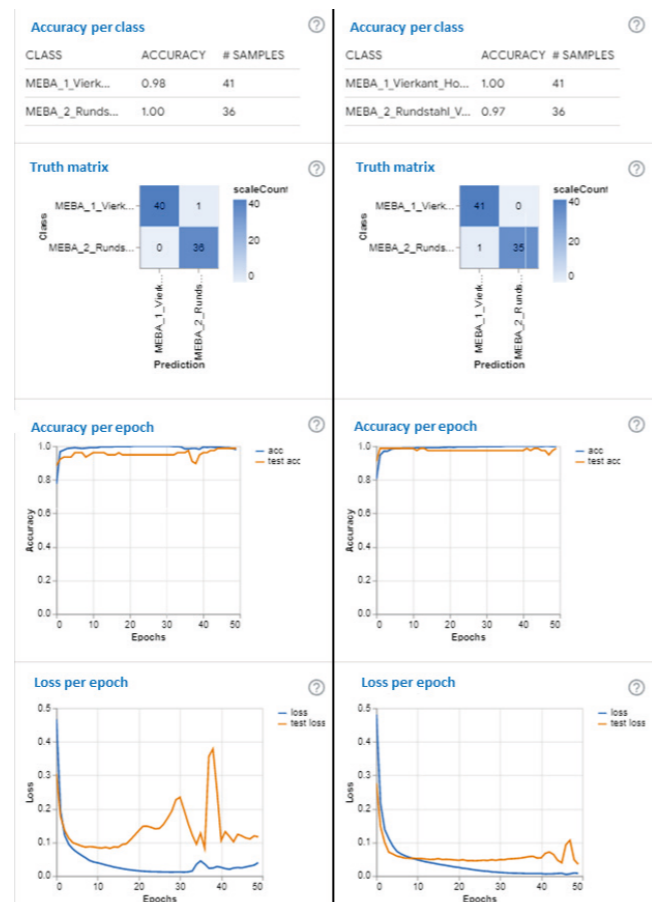


Figure 5: Statistics of training 1 and 2.

Fig. 6 shows the results of training 3. The „loss“ curve already drops sharply after a few epochs and remains at a low level, except for an increase in epochs 15 to 22. The accuracy „acc“ also compares at very high values over the entire course, except for a small drop also in epochs 15 to approx. 22. The resulting accuracy per class between 98 and 99.99 % and the confusion matrix with only one wrongly assigned data point indicate a very good classification ability of the AI model.



Figure 6: Statistics for training 3.

The AI model was transferred to the NodeOne in Keras h5 format, and the program flow at live classification was loaded. Data acquisition for live classification was performed under the same conditions as for dataset creation. Fig. 7 shows the results of the validation of the experiment. The upper plot shows the time series of the measurement signal to the respective class. Some peaks can be seen in the measurement signal, but they have little effect on the classification because they do not have a strong influence after feature extraction. The outliers were probably caused by a defect in the cabling; however, this could not be clarified conclusively.

The lower plots are the probabilities or accuracies from the AI model. The scales are set differently so that something can be read in both progressions. The mean value „Mean“ is also noted under the courses, which gives an accuracy between 99.84 % and 99.95 % for the „gouraud“ model.

4 DISCUSSION AND CONCLUSIONS

The experiment was aimed to create a reliable detection system for numerous machine and tool states using vibration sensor data with deep learning methods. A pre-

trained convolutional neuronal network was used that has been trained on a huge dataset and comes with good performing feature layers within the model. Using transfer learning, it was fitted to work with new data created within the experiment. An accelerometer to record the vibration of different machine states was used, e.g., such as an unbalanced and healthy engine, or different tool conditions, such as worn out or usable. In this experiment the focus was on the detection of geometric properties of the workpiece while it is getting cut by an industrial band saw machine. There were two defined states, that the classification was supposed to differentiate:

- Class 1: hollow metal workpiece with square dimensions
- Class 2: full metal workpiece with round dimensions

The broader goal is to create an AI model that can detect various inhomogeneities in workpieces such as unwanted inclusions or blow holes that could lead to miss matched machine operation

settings and ultimately in bad product quality and faster that necessary wear of the machine and tools itself. The trained AI-model from the experiment was able to generate good accuracies and has also shown in live operation on the Node-One that they also present good classification accuracies in the application. The representation settings during feature extraction have a large impact on the quality of the generated datasets. It has been shown that the „gouraud“ shading setting for da-

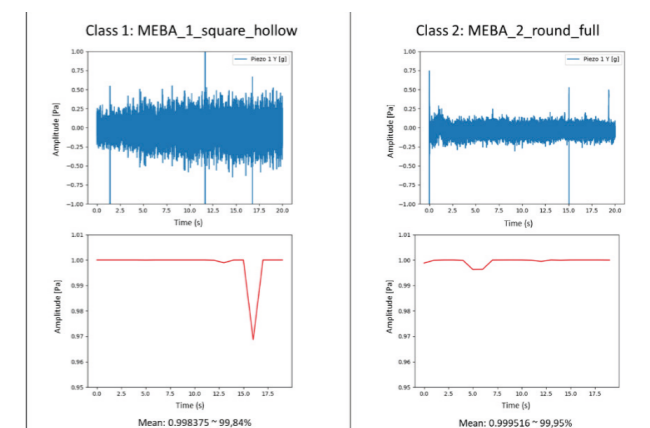


Figure 7: Results of the live monitoring.

datasets with a smaller dimension than the expected input format of the AI model led to good results in classification ability. The model has already achieved good results despite small datasets, showing that the applicability of the presented learning process is suitable for rapid prototyping projects.

Despite the good result in the experiment, it should be noted that it cannot be ruled out that the model is still in the overfitting range. This is indicated by the increase of the orange curve of „test loss“ over the number of epochs. The model is most likely not able to recognize the geometry with different materials or workpiece dimensions. A generalized model requires the acquisition of a much larger data set with more variance in workpieces. In addition, the condition and properties of the saw blade, e.g., wear progress, and information about the sawing process, e.g., process parameters from the machine control system, still need to be considered and included in the data sets.

ACKNOWLEDGEMENT

The research project IGF 22150 N „Analyse der Anwendung von Machine Learning Methoden für die Echtzeitüberwachung der Zerspanung inhomogener Werkstoffe zur Beurteilung von Werkzeugzustand und Werkstückqualität in kleinen und mittelständischen Unternehmen“ from the Research Association Forschungsgemeinschaft Werkzeuge und Werkstoffe e.V. (FGW), Papenberger Straße 49, 42859 Remscheid, Germany, is supported by the Federal Ministry of Economic Affairs and Climate Action through the German Federation of Industrial Research Associations (AiF) as part of the program for promoting industrial cooperative research (IGF) based on a decision by the German Bundestag.

REFERENCES

1. Statista GmbH, Künstliche Intelligenz rechnet sich, <https://de.statista.com/infogra-fik/16992/umsatz-der-in-deutschland-durch-ki-anwendungen-beeinflusst-wird/>, last accessed 2023/07/12
2. Teti, R., Jemielniak, K., O'Donnell, G., Dornfeld, D.: Advanced monitoring of machining operations. *CIRP Annals* 59(2), 717–739 (2010)
3. Lauro, C. H., Brandão, L. C., Baldo, D., Reis, R. A., Davim, J. P.: Monitoring and processing signal applied in machining processes – A review. *Measurement* 58, 73–86 (2014)
4. Stavropoulos, P., Papacharalampopoulos, A., Vasilidis, E., Chryssolouris, G.: Tool wear predictability estimation in milling based on multi-sensorial data. *The International Journal of Advanced Manufacturing Technology* 82, 509–521 (2015)
5. Koike, R., Ohnishi, K., Aoyama, T.: A sensorless approach for tool fracture detection in milling by integrating multi-axial servo information. *CIRP Annals* 65(1), 385–388 (2016)
6. PCB Synotech GmbH, <https://www.pcbpiezotronics.de/>, last accessed 2023/07/12
7. Kneifel, J., Roj, R., Woyand, H. B., Theiß, R., Dültgen, P.: An IIoT-Device for Acquisition and Analysis of High-Frequency Data Processed by Artificial Intelligence. *IoT* 4(3), 244–264 (2023)
8. MEBA Metall-Bandsägemaschinen GmbH, <https://meba-saw.com/>, last accessed 2023/07/12
9. GoogleCreativeLab - Teachable Machine, <https://github.com/googlecreativelab/teachablemachine-community>, last accessed 2023/07/13
10. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv:1704.04861* (2017)

AI-based Wear Analysis of Sawing Processes using Machine Data

Jens Kneifel¹, Christian Steinberg¹, Christoph Böttinger¹, Robin Roj¹, Ralf Theiß¹, Peter Dültgen¹, Stefan Bracke², Jörg Blecher³, Norbert Böhme⁴, and Christian Klostermann⁵

- ¹ Forschungsgemeinschaft Werkzeuge und Werkstoffe e. V., Papenberger Str. 49, 42859 Remscheid,
² Lehrstuhl für Zuverlässigkeitstechnik und Risikoanalytik, Bergische Universität Wuppertal, Gaußstraße 20, 42119 Wuppertal
³ August Blecher GmbH & Co KG, Wuestenhagener Str. 65, 42855 Remscheid
⁴ Böhme und Weihs Systemtechnik GmbH & Co KG, Linderhauser Str. 153, 42279 Wuppertal
⁵ Klostermann Ingenieurbüro und Vertriebsgesellschaft mbH, An der Hasenjagd 5, 42897 Remscheid

Abstract. In the sawing industry, accurate assessment of wear in saw blades is crucial for optimizing maintenance and replacement strategies. This paper presents an AI-based approach for wear analysis in sawing processes using acoustics, vibration data, and power consumption metrics. By analyzing large amounts of machining data, including vibration signals and acoustic emissions, the conditions of saw blades can be predicted, enabling proactive maintenance and cost savings.

The proposed method also allows for the optimization of saw cuts by determining the optimal cutting speed, feed rate, and cooling for different materials, leading to improved quality and performance. The study combines various data analysis techniques, including machine learning algorithms and statistical methods, to develop predictive models for estimating saw blade life and optimizing cutting processes.

The wear condition of the saw blade is determined based on parameters such as vibration, acoustic signals, power consumption, temperature, humidity, and properties of generated chips. The process and quality data are stored in a database, and implicit correlations are identified to develop a prognosis model. The presented methodology provides a framework for implementing predictive maintenance and quality optimization in the sawing industry, enabling companies to minimize downtime, plan operations more efficiently, and achieve cost savings. Further research and development in this area are expected to yield continuous improvements in efficiency and performance.

Keywords: Predictive Maintenance, Predictive Quality, Process Data, Sawing Processes, Sensor Data Analysis.

1 INTRODUCTION

In the sawing industry, wear is assessed by experienced operators based on acoustic emissions, but this is not a reliable method for realizing the full potential of the saw blade. Often, experienced operators cannot recognize the point at which saw teeth simply need to be resharpened before they break or are otherwise damaged and need to be replaced. Resharpening saw teeth can save significant costs of up to 70 % compared to replacing individual saw blade teeth. The cost savings result from simpler, partially automated processes, shorter processing time and the elimination of material costs. Currently, there are separate processes in industry: The manufacturing company owns the saw blade and uses it until it needs to be repaired. It then hires a professional to maintain the saw blade and returns it until it needs to be repaired again. By looking at the wear values using artificial intelligence (AI), the repairer can get an idea of how many saw blades need to be resharpened within a certain period of time to prepare for absence and organize replacement.

The implementation of machine learning in the saw blade industry also offers the possibility of identifying patterns in wear behavior and deriving predictive models from them (predictive maintenance). These models enable accurate estimation of saw blade life in different application scenarios, leading to more efficient resource use and cost savings. In addition to predicting the wear conditions, AI also allows optimization of the saw cuts themselves. By analyzing large amounts of sawing process data, the optimal cutting speed, feed rate and cooling for different materials can be

determined (predictive quality). These insights can help improving the sawing performance, minimize wear, and ensure highly accurate cuts.

The integration of AI and data analytics into the saw blade industry thus promises more efficient use of resources, improved productivity, and enhanced quality of sawing processes. Companies can better plan their operations and minimize downtime while increasing customer satisfaction. The ongoing development in this area ensures further improvements in efficiency and performance.

2 STATE OF THE ART

The wear condition of a saw blade can be determined in various ways. Traditionally, this is done in small and medium enterprises (SME) by visual inspection by a trained employee. Manual inspection means that it takes several hours to recondition a saw blade. Also, the human factor must be taken into account, so it is possible that defects are misidentified or even overlooked. Recently, Bankar et al. qualified an audio-based method for detecting the sharpness of saw blades, this is supported by AI, which trains models that perform the evaluation of the acoustic signals [1]. In the same year, Zhuo et al. presented similar results; by analyzing acoustic emissions during the sawing process, the wear condition of band saw blades could be determined in real time with an accuracy of 97.71 % [2].

A study from 2021 combined various sensors in the wood industry to monitor tool wear. The sensor measurement data was analyzed by a neural network and was able to determine wear with 92 % accuracy [3]. By attaching several sensors, this method is suitable as a retrofit variant in sawing operations. Furthermore, to improve the efficiency of saw blade reworking, automated image evaluation of the individual saw teeth by image processing algorithms is suitable. This offers higher efficiency, higher precision, and lower costs through more efficient use of personnel. The tooth pitch, the tooth root depth, the radius of curvature of the tooth form, and the clearance angle are measured and then compared with ideal values [4]. The content described below combines research results into an information network. In the sawing operation,

the saw blades can be replaced in time, so that the saw blade teeth only need to be resharpened instead of being replaced. This leads to financial savings of up to 70 %.

3 MATERIALS AND METHODS

A database was created with all process data and quality data of the saw blade in order to identify implicit correlations and to develop a new prognosis model based on the database. In creating the parameter matrix, the five categories of saw blade data, material data, machine data, cutting data, and application data are distinguished (cf. Fig. 1).

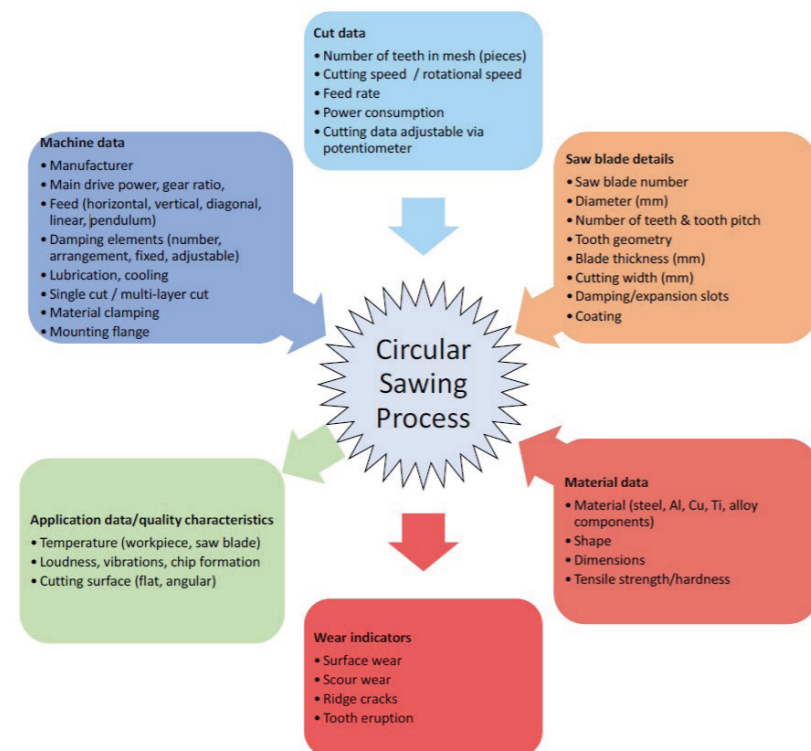


Fig 1. Parameters influencing the sawing process.

In the evaluation of saw blade wear, cutting edge wear, tooth fractures, cracks, damage to the blade flank, scour wear, and comb cracks were examined. The most important characteristics that indicate the wear condition of the saw blade teeth are vibration of the blade, acoustics of the cutting process, power consumption, temperature, humidity in the machine housing, and properties of the generated chips. Chips, vibration, and power consumption were considered for wear condition measurement.

The machining of stainless steel and mild steel produces

chips of different colors, reddish for stainless steel, and dark blue for mild steel. Frayed edges, increased roughness of the chips, and small particles on the chip surfaces are meaningful indicators of advanced wear of the saw blade. Curvatures, length, and kinks of the chips are not indicators of wear but can indicate the cutting position.

As the saw teeth wear, dulling occurs, generating frictional energy at the cutting edge and creating higher resistance. This resistance occurs in the form of thermal energy, which heats the saw blade, causing the blade to lose stiffness and begin to vibrate. For optimal measure of the vibrations, sensors were placed inside and outside the machine housing. Two sensors were installed in the machine: one on the motor and one near the workpiece being machined. In addition, an acceleration sensor was mounted on the machine housing.

The use of a saw blade causes vibrations due to the repetitive engagement of the saw teeth and the overcoming of counterforces of the workpiece. The saw's motor control regulates the current to maintain the required rotation speed, even with varying forces. To perform these measurements, a sensor network was developed to collect all relevant data and allow the extraction of characteristics. Vibrations and structure-borne noise were measured at two locations on the sawing machine using piezo accelerometers. A microphone recorded the emitted room sound. In addition, voltage and current were measured directly at the saw's converter in the control cabinet. All signals were recorded synchronized in time in a measuring system.

For a meaningful analysis of the process, a reduced number of sensors was sufficient to make the measurement setup scalable and cost-efficient. At the beginning of the saw blade life, low vibrations and power consumptions could be measured. With increasing wear, both increased as it can be seen in Fig. 2. In the next steps this information was interpreted by an AI model.

Various methods could be used to analyze the degradation behavior of saw blades. A convolutional neural network detects the wear of individual teeth by training it with data from intact and worn teeth. Piezo accelerometers on the saw can be used to analyze degradation behavior during the sawing process, since saw blade wear increases vibrations and power consumption on the saw. The analysis of the power consumption, which is constantly adjusted to ensure consistent sawing properties and thus counteract wear, was also a goal.

Cutting tool wear is generally caused by high mechani-

cal forces and the associated high temperatures. These loads depend on various parameters, such as the material of the cutting tool, the material of the workpiece, the cutting speed, the feed rate, the cooling, and the environmental conditions. To assess the reliability of the cutting tools, images of the teeth at different stages were analyzed. This enables an estimation of the degree of wear and the remaining service life, depending on the material being cut. By statistically evaluating the color gradient of the images, conclusions could be drawn about the wear type and defects of the cutting tool. The tool geometry was captured from different perspectives, and characteristic values were used to develop a simple evaluation method for tool wear. This contributes to image-based analysis of machining processes and the wear behavior of cutting tools, with statistical methods enabling comparability.

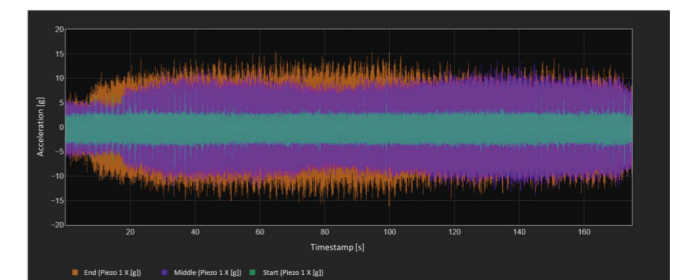


Fig. 2. Acceleration measurement of a saw blade at the beginning (green), in the middle (purple) and at the end of its life (orange).

With increasing wear, acceleration sensors record stronger vibrations on the drive shaft and the housing of the circular saw during the sawing process. Long-term evaluations of this data allow conclusions to be drawn about service life and wear depending on the material. The recorded measured values were processed into characteristic values, which were evaluated by AI.

The evaluation of the sawing process parameters for the wear progress was carried out by an overall view of the operating process parameters vibration, power consumption, and noise. The noise sensor allowed conclusions to be drawn about significant influences of the sawing process. However, the noise pickup was influenced by background noise and other interfering noises, which leads to uncertainties in the filtered signals.

The vibration parameter seemed to be more suitable since it was measured directly at the machine and thus represents the frequency spectrum of the machine. However, the measurement of the power consumption allowed a more precise representation of the wear.

The parameter of power consumption was suitable for

continuously describing and predicting the progress of wear. Vibration and noise development allowed systematic deviations from the ideal process to be detected but are not evaluated as having a high resolution. To save resources, saw blades were reworked several times, with the diameter of the saw blade continuously decreasing. A reworked saw blade did not show significantly different performance to the first use. Center of gravity and dispersion with respect to the measurement series of the process key indicator "cut surface" showed no differences.

In addition, various process parameters were investigated using AI methods. „Decision Trees“ served as a basic algorithm for forecasting purposes and „Random Forest“ for prioritizing the most targetable parameter.

A virtual CAQ (computer aided quality) server was used for the research (cf. Fig. 3.). The saw blade characteristics and parameters were important for the operation of the server. By reinstalling a CAQ server, it was possible to operate this server directly in a cloud.

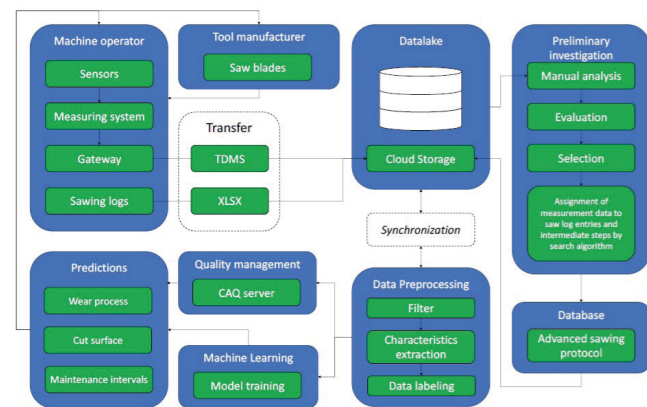


Fig. 3. Data flow in the cloud infrastructure.

In order to assign the collected measurement data to the respective correlating context from the saw logs, a mixture of manual sifting and programming of a search algorithm was necessary. The assignment was recorded in a central database, which was accessed for further analysis. This ensured that different evaluators refer to the same data sets and thus the results are comparable. The data collected at the machine was stored in the binary TDMS file format. Handwritten saw logs were recorded in cloud storage as an Excel spreadsheet. Connected to the cloud, in addition to the desktop client for pure data synchronization, are Python libraries that have been adapted to provide access to the respective files to be processed within a data pipeline. After almost every processing step, the result was uploaded back to the cloud

storage so that all evaluators received the same data status.

For the development of prognosis models regarding tool life, the sawn material, the cutting speed, and the feed rate are decisive (cf. Fig. 4). The prognosis tool combines measured values with handwritten sawing logs. During live monitoring, only the area in which the saw blade is fully engaged was considered. For this purpose, a search algorithm analyzes the data from the power measurement to determine the point in time when most of the teeth are in mesh. To enable reliable predictions, the underlying models must be robust and trained with extensive data sets.

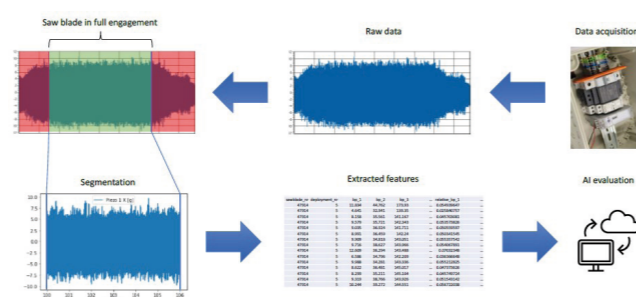


Fig. 4. Perform feature extraction from subsets via batch processing.

Real-time monitoring requires minimal time lag between data collection and model prediction. Recorded measurement data could be examined as a whole or in small segments. The examination of the entire time series was useful for the analytical evaluation and the derivation of knowledge about the saw blade used. The evaluation of small segments is purposeful for the detection of anomalies and progressive wear mechanisms during the sawing process.

The aim was to perform short-time Fourier transforms, in the form of a spectrogram. The time interval was chosen such that a subset contains excitations from all saw teeth at least three times. By integrating preprocessing into the feature extraction pipeline, the efficiency of the AI training process could be increased. In addition, data transfer and storage were accelerated because the generated subsets produce smaller files. During processing, it was recognized that software-based tools were needed to provide user-friendly access to non-specialist operators. As a result, three dashboards were developed with a strong emphasis on simple and intuitive operation.

A data analyzer tool developed outside the project was used to evaluate the measurement data. It provides a comprehensive overview of the measurement data and

allows manual individualization of the corresponding diagrams, so that the diagram of interest can be configured and displayed in a targeted manner. The tool combines data from different servers, which can be selected accordingly. The menu navigation allows the selection of various preview elements, giving the user an overview of the available analysis values in a short time. Furthermore, the display of the diagrams, selection, and zoom functions were stored in order to analyze specific areas. In addition to the real-time processing of the sensor data, the power consumption was collected as a timeseries and analyzed on a per cut basis. Therefore, features were extracted specifically for power consumption characteristics and stored in a separate dataset in which each data-point represents a different state of wear. The power consumption data set was extracted by committee member LZR (Chair of Reliability Engineering and Risk Analysis, University of Wuppertal).

The dashboards were used to prepare for AI-based further processing. Through test runs, the AI could be understood and adjusted in order to prepare a better differentiation of the data. After analysis, feature extraction jobs could be sent to the backend tool running on a powerful separate server. The metrics derived from the power consumption were processed by a feature extraction program, while another AI application ran training algorithms and compared them in accuracy. The best model was automatically recognized and exported.

4 RESULTS

In the next step, the model training could be started immediately without any further preparatory steps. First, the target value was selected from the imported data sets. Selection algorithms were then triggered in the background, which extracted the corresponding training parameters and display them to the user before the start of the training phase. Since it was not clear at the beginning of the training phase which AI algorithms provide the most meaningful results, all stored models were tested and evaluated by means of error rate. Out of 19 deposited models, the „Extra Trees Regressor“ provided the best results. For an optimal evaluation of the performance of all deposited models, the most important key performance indicators were displayed to the user and the final model could be exported automatically for further use.

The amplitudes of the sound pressure signals increased with wear, especially at the end of tool life. Key indicators were extracted for each cut, such as break points, total process time, and various statistical values of the sound pressure level. These key indicators allowed a better observation of the wear behavior of the sawing tools. Due to the grinding-in process, a decreasing trend could be seen at the beginning. Towards the end of the product life, an increased scatter of the measured values could be noticed before the sound pressure level increased significantly, thus indicating advanced tool wear. A simple user interface informs the machine operator of the saw blade's condition and estimated remaining life. The values displayed are to be initially reduced to the AI predictions, including remaining cut area, number of cuts, and warnings for immediate tool replacement.

A dashboard is used for monitoring and visualization. It provides an important overview of the circular saw's relevant metrics and output. This overview focuses primarily on performance, profitability, and efficiency. Process optimizations could be derived by comparing the collected data with internal processes. There are correlations between employees, layers, materials, and wear. These characteristic values allow specific conclusions to be drawn about the overriding profitability and support decisionmaking processes. Key figures were extracted from the recorded signals and an AI model was applied. The intersection number was used as the target value. Thus, for each measurement, a known cut number was available (actual value) as well as a predicted cut number from the AI model (regression). The two values were compared to assess the wear progress. If the actual value is higher than the prediction, the wear is low. However, if the actual value is lower than the prediction, the wear is more advanced than expected and the operator may be warned. Additionally, the values are considered in relation to the target final cut number to indicate a percentage saw blade degradation. Fig. 5 shows that the 19 best models were identified from the training of 25 algorithms. The model „Extra Trees Regressor“ achieved the lowest mean absolute error.

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
et	Extra Trees Regressor	7.2208	178.8096	12.7731	0.7781	0.5934	0.4368	2.7700
catboost	CatBoost Regressor	8.3245	192.8559	13.3875	0.7606	0.5818	0.5196	29.2720
rf	Random Forest Regressor	9.5028	211.4251	14.1838	0.7372	0.5996	0.5921	7.0180
lightgbm	Light Gradient Boosting Machine	9.2451	215.4099	14.2226	0.7329	0.5888	0.5433	2.7660
gbr	Gradient Boosting Regressor	10.1071	221.6970	14.4948	0.7254	0.6121	0.6433	5.5300
xgboost	Extreme Gradient Boosting	9.4855	224.1530	14.5585	0.7213	0.6186	0.5575	4.0310
ada	AdaBoost Regressor	15.3611	388.9005	19.6080	0.5155	0.7704	1.0558	2.0970
dt	Decision Tree Regressor	12.4037	444.4211	20.7760	0.4465	0.7718	0.6799	1.7810
knn	K Neighbors Regressor	18.2597	534.3534	23.0346	0.3304	0.7990	1.0423	1.7060
dummy	Dummy Regressor	23.7194	808.1040	28.3971	-0.0097	0.9684	1.7677	2.2430
huber	Huber Regressor	21.2293	1875.2964	33.3904	-1.3026	0.8652	1.3407	1.7910
par	Passive Aggressive Regressor	38.0897	2507.9522	48.1757	-2.1604	1.4994	1.7005	1.6740
omp	Orthogonal Matching Pursuit	18.5380	2667.7546	38.3301	-2.2010	0.8112	1.1560	1.6210
br	Bayesian Ridge	24.8942	5375.6009	58.9125	-5.6996	0.8777	1.3164	1.6830
lasso	Lasso Regression	20.5904	7254.7776	59.0399	-8.0764	0.8066	1.2248	1.9850
llar	Lasso Least Angle Regression	20.1683	7490.7837	52.8168	-8.1092	0.8098	1.2709	1.6370
en	Elastic Net	20.5157	7517.0476	58.6670	-8.4129	0.8110	1.2251	1.6810
lr	Linear Regression	27.2939	51600.0122	126.1334	-60.5443	0.8243	1.5118	1.8690
ridge	Ridge Regression	63.5040	1815263.4697	488.9516	-2134.6673	0.8364	3.0993	1.6220

Fig. 5. Model comparison with listing of training metrics

5 DISCUSSION AND CONCLUSION

By analyzing the „Feature Importance“, it was possible to determine which metrics had the strongest influence on the predictions of the AI model (cf. Fig. 6). The first metrics had a significant impact, while the following ones were increasingly redundant and could potentially be omitted without sacrificing prediction accuracy. This allowed a reduction in the data set and reduced the influence of overfitting. The „Extra Trees Regressor“ achieved a high training score close to 1, indicating high precision. However, the larger gap between the training score and the crossvalidation score indicates possible overfitting problems. Nevertheless, the model achieved a good average validation score of 0.8, suggesting sufficiently accurate results.

The model was trained on known data and then validated on unknown data. During training, the model showed a high accuracy of 0.98 (training score). When applied to unknown data, the validation score varied between 0.62 and 0.92 with an average of 0.78. This shows that the accuracy of the model was lower on validation data than on training data, and such reduced accuracy was to be expected. Optimizing the hyperparameters and reducing overfitting could improve the model. It was applicable, but for comprehensive generalization, it

needs further training. This can be done to continuously optimize the model and add new data.

Another indication of the accuracy of the model is the plot of the prediction error. The R² value of the „Extra Trees Regressor“ model is quite high at 0.817, indicating that the model provides good predictions for the observed data. Most of the data points on the plot are along the zero line, which represents the average error. Only a few points deviate greatly from it, suggesting that the model is applicable in most measurements. Improvements to the data set and an increase in the amount of training data may lead to further improvements here.

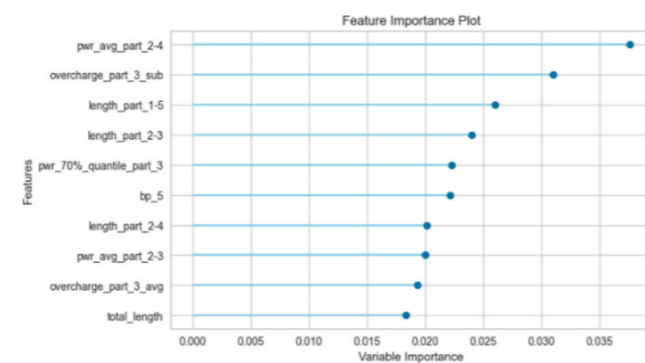


Fig. 6. Relevance of data within model as Feature Importance Plot

In the t-SNE Manifold Plot (cf. Fig. 7), high-dimensional data sets were analyzed with the target „cutting number“, color-coded, and plotted on a scale from 0-120. The plot shows that data points with similar cutting numbers (points with similar colors) are grouped in similar areas. This close spatial proximity between points in the plot indicates a strong correlation between the features and the target values (predictions) and confirms the information sought about the wear condition in the data.

The anomalies located in the t-SNE Manifold Plot can be identified by color differences and deviations from the general pattern. These can attribute to various reasons, such as a too short learning phase of the AI, incomplete data sets, or incorrect measurement data, on which the AI was trained. Due to the limited impact of the anomalies, they have no significant relevance for the accuracy of the prediction model since the overall picture is consistent and the correlation between the data points remains unambiguous.

By integrating t-SNE Manifold Plots into AI-driven wear analysis in the saw blade industry, the accuracy of predictions can be determined, enabling the planning of preventive maintenance measures. The gained insights contribute to fully harnessing the potential of the saw blades and optimizing operational processes. The ongoing development and application of t-SNE Manifold Plots and AI techniques thus offer a promising future for the saw blade industry and further improvements in efficiency and performance resulting in increased market competitiveness.

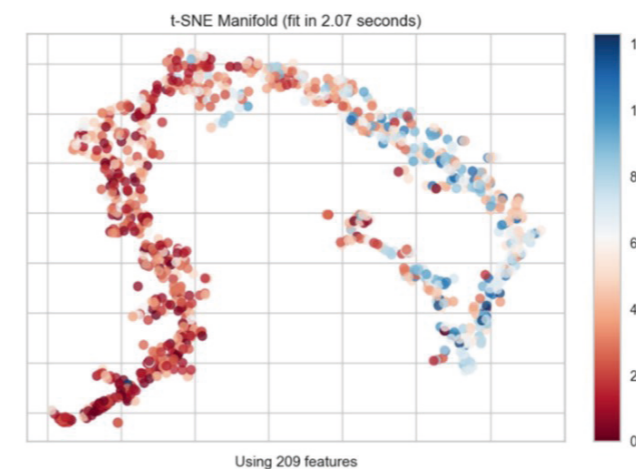


Fig. 7. Similarity of data points in t-SNE Manifold Plot.

ACKNOWLEDGEMENT

The research project EFRE-0801649 „SmartCut - KI-basiertes System zur Steigerung der Effizienz und Nachhaltigkeit von Sägeprozessen“ is supported by the Ministry of Economic Affairs, Industry, Climate Action and Energy of the State of North Rhine-Westphalia, Germany. The authors thank the European Regional Development Fund and the Project Management Jülich | Forschungszentrum Jülich GmbH.

REFERENCES

1. Bankar, P., Kakade, S. M., Shukla, P., Gupta, A., Bagewadi, M., Kulkarni, R. S.: Audio Based Detection of Saw Blade Sharpness Using Machine Learning. In: International Conference on Signal and Information Processing 2022, IConSIP, vol. 2, pp. 1–5. IEEE, Pune, India (2022)
2. Zhuo, R., Deng, Z., Chen, B., Liu, T., Ge, J, Liu, G., Bi, S.: Research on online intelligent monitoring system of band saw blade wear status based on multi-feature fusion of acoustic emission signals. The International Journal of Advanced Manufacturing Technology 121(7-8), 4533–4548 (2022)
3. Nasir, V., Dibaji, S., Alaswad, K., Cool, J.: Tool wear monitoring by ensemble learning and sensor fusion using power, sound, vibration, and AE signals. Manufacturing Letters 30, 32–38 (2021)
4. Wang, Y.-C., Lin, J.-C., Chiu, S.-F.: Automatic inspection system for dimensional measurements of the saw blade milling cutter. Journal of Industrial Engineering and Management 3(2), 370–382 (2010)

TIADE: A Toolbox for Identifying and Analyzing Anomaly Detection Errors

Marvin Grimm¹ and Erik Rodner¹

¹KI-Werkstatt/Fachbereich 2,
University of Applied Sciences Berlin,
Wilhelminenhofstr. 75A, 12459 Berlin, Germany
marvin.grimm@htw-berlin.de

Abstract. Analyzing the errors and behaviour of machine learning algorithms is a key task. However, complex results are often only boiled down to single measures unable to reveal important systematic differences. While many evaluation frameworks have been proposed in the computer vision domain, benchmarking anomaly detection algorithms for time-series is still unexplored. Therefore, we developed TIADE: a Toolbox for Identifying Anomaly Detection Errors. TIADE is a framework and graphical user interface for the performance analysis of anomaly detection algorithms for multi-variate time-series. It allows for in-depth error analysis by differentiating between multiple error types, such as localization inaccuracies, false positives, and systematic misses. Therefore, algorithm evaluation with TIADE goes beyond single performance values, such as the area under the ROC curve. We further demonstrate the advantages of TIADE on challenging real-world data. The code for this project will be available on GitHub.

Keywords: Error Diagnosis · Evaluation · Anomaly Detection · Benchmarking · RobustML.

1 INTRODUCTION

Anomaly detection on spatio-temporal data has been an active area of research due to its importance in various domains. However, there is a lack of benchmarking tools for anomaly detection algorithms that detect anomalous intervals [3] rather than extreme points. Without proper benchmarking tools and due to the typical academic tendency in our field of focusing on “beating the previous state-of-the-art”, algorithmic results are often boiled down to single performance measures, such as the area under the receiver operator curve or mean average precision [3]. However, this leads to a biased view on algorithm performance and likely hides significant differences in the types of errors the algorithms make. The same conclusions have been made in the computer vision field, leading to benchmarking frameworks such as TIDE [4] and TIVE [6] or analysis papers such as [9] for 3d object detection.

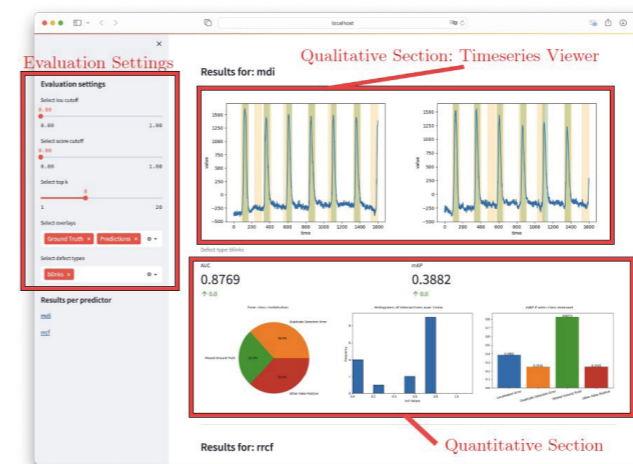


Fig. 1. Anomaly detection performance analysis with TIADE using its simple one-page application

In this paper, we therefore propose the first benchmarking framework that allows for a comprehensive evaluation of anomaly detection algorithms. Our framework incorporates both quantitative and qualitative evaluation metrics (further explained in Sect. 3), taking into account the temporal aspects of the data. We demonstrate the effectiveness of our framework through experiments on realworld EEG data (Sect. 4). TIADE can be easily adapted to new datasets and algorithms and has the potential to provide insights into the strengths and limitations of current state-of-the-art anomaly detection algorithms.

2 RELATED WORK

Anomaly detection in times series Detecting anomalies in time-series is a standard problem in machine learning and signal processing. However, one has to differentiate previous work with respect to the type of anomaly under consideration. Figure 2 illustrates the most common anomaly definitions for uni-variate time-series. Whereas the majority of papers [10, 7] focuses on detecting single extreme points, other papers try to detect a change of the trend in a time-series [1], which is especially important for event detection and automatically detecting critical changes of system usage [12]. The focus of this paper is on detecting extreme intervals in online or even offline settings [3, 11]. Algorithms for extreme interval detection have shown to be a great tool for scientific exploration [2].

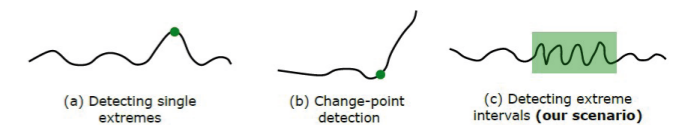


Fig. 2. Different types of anomaly detection errors tackled in previous works.

Benchmarking beyond single performance measures

One of the first papers arguing for a more detailed experimental benchmarking beyond single values is the paper of Hoiem et al. [5]. Their work systematically listed the different types of object detection errors that can occur, such as confusion with dissimilar or similar objects or localization errors. Although our work is focusing on (multivariate) time-series data instead of images, the nature of our anomaly detection problem (detecting extreme intervals) can be regarded as an object detection problem in 1D. Furthermore, our scenario is even simpler, since unsupervised anomaly detection in our case does not focus on categorizing detected anomalies. The latter could however be integrated in future work. Our approach draws a lot of inspiration from TIDE [4], which extends [5] by e.g. analyzing all detections rather than the top ones. Extending these results to the spatiotemporal domain is done in [6] for videos. Another line of work that should be mentioned here concentrates on analyzing algorithm biases [13].

3 ANALYZING AND CATEGORIZING ANOMALY DETECTION ERRORS

Inspired by the TIDE framework [4] for analyzing object detection results, we categorize errors made by an algorithm into several error types that are later on used in an error attribution scheme.

3.1 ERROR TYPE DEFINITIONS

Since our focus is unsupervised anomaly detection without classification, our error types reduce to four different categories: Location Error, Duplicate Detection Error, Missed Ground Truth (GT), and False Positive. Please note that our anomaly detections are indeed time intervals and not single points. Comparing a detection interval $A = (a1, a2)$ to a ground-truth interval $B = (b1, b2)$ is therefore made with the intersection-over-union metric: $IoU(A,B) = (A \cap B)/(A \cup B)$.

Localisation Error occurs when the detected anomalous interval has a different spatial location compared to the GT anomalous interval but is close-by. This error type

captures the spatial accuracy of the anomaly detection algorithm and measures its ability to correctly identify the location of anomalous intervals in the spatio-temporal data. The error-type is affected by a user-defined threshold for the Intersection over Union (IoU): If an anomalous region is partially detected ($\text{IoU} > 0$), but the IoU is below a chosen threshold T , the detection will not be classified as a false positive, but as a localisation error. The threshold is highly application-specific and can therefore be adjusted in TIADE by the user (cf. Fig. 1).

Duplicate Detection Error occurs when an anomalous interval is detected multiple times by the algorithm. The error-type measures the for example effectiveness of any non-maximum suppression performed as a post-processing by the algorithm.

Missed Ground Truth or False Negative occurs when the algorithm completely fails to detect an anomalous interval that is present in the ground truth. This error type captures the sensitivity of the anomaly detection algorithm and measures its ability to detect all anomalous intervals in the data.

Other False Positives occur when the algorithm detects anomalous intervals that are not present in the ground truth, and are neither a duplicate detection nor a localisation error. This error type captures the specificity of the anomaly detection algorithm and measures its ability to accurately identify only true anomalous intervals.

Overall, the incorporation of multiple error types allows us to evaluate the performance of anomaly detection algorithms on anomalous intervals in a more nuanced and comprehensive manner.

3.2 ESTIMATING THE POTENTIAL OF AN ALGORITHM WITH ORACLE PERFORMANCE

A standard metric that is widely used for object detection and has been adapted for anomaly detection [3] is the mean average precision (mAP) score. By measuring the area under the recall-precision score, the metric is independent of any threshold applied to the anomaly detection scores of the algorithm. We can therefore perform performance analysis independent of the sensitivity requirements of the application. By measuring the contribution of these error types to the mean average precision

(mAP) score using an oracle, we can further quantify the impact of each error type on the overall performance of the anomaly detection algorithm. The oracle replaces all predictions of one error-type with perfect predictions. The resulting mAP is then visualized with respect to the error-type that has been removed. This enables us to gain a comprehensive understanding of the strengths and limitations and potential of the algorithm

3.3 ANALYSIS IN TIADE

The TIADE user interface is comprised of the following regions:

Evaluation Settings To incorporate a users possible domain knowledge into the evaluation process, we added the following adjustable parameters:



Fig. 3. (Left) The settings panel of TIADE (Right). We can see both the selected top-k and score-cutoff being applied to the time series viewer (red). Further, only the selected defect type (blinks) is considered.

- IoU Threshold:** This setting is applied to the Time Series Viewer, as well as the Error Class Contribution (see Fig. 1). Predictions that have an IoU value below the set threshold T will be omitted from the Time Series Viewer. The threshold is further used to distinguish between error classes Other False Positive and Missed GT (see Sect. 3.1).
- Confidence Score Threshold:** This setting is applied as a filter to the predictions made by the algorithm. Therefore, every qualitative metric as well as the Time Series Viewer are affected by this setting. It can be used to determine a sensible anomaly classification threshold for the algorithm. For threshold-independent comparisons, it is reasonable to leave the value at 0.0, *i.e.* all predictions are considered.

- Top-K:** This setting can be used to incorporate domain knowledge into the evaluation. For instance, if the user knows that there will be only one anomalous interval in the time series, using only the prediction for which the algorithms have the highest confidence might lead to more comparable results. The user is advised to make sure that the detected anomalous intervals are distinguishable by their score, as the selection of the top k detections might not be deterministic if multiple detected regions share a confidence score.
- Selected Overlays:** This gives a user the possibility to either show or hide the detected region(s) and ground truth(s). This is especially useful for long time-series with many defects.
- Selected Defect Types:** This gives a user the ability to limit the defect types considered when evaluating the algorithms. This setting affects every metric.

Qualitative Analysis To support users of our framework with a qualitative analysis, we provide a way to visually inspect the output of the algorithm using a time-series viewer. The interface allows selection of defect types to be displayed, as well as adjustable IoU and score cutoffs.

Since the dimensionality of data point in the time-series can be high, we need to reduce the visualization of the time-series to the most important parts. Selecting only relevant variables of the data points could be one option, however, some variables that have been used to detect an anomaly could be missed in the visualization. Therefore for opted for an dimensionality reduction in each data point. In particular, we use principal component analysis (PCA) applied to all data points individually and reduce the dimension to 1, *i.e.* transforming the multi-variate time series to a univariate one. The user is able to overlay the ground-truth, as well as predictions made by the individual algorithms with an adjustable IoU-cutoff, adjustable confidence score cutoff, as well as an adjustable top-K. The viewer enables users to identify cases where the algorithm consistently fails to detect anomalous intervals or generates false positives. The adjustable cutoff allows users to arrive at a meaningful score threshold for binary classification based on visual inspection using production data. The IoU cutoff setting enables the user to emphasize an algorithms localization ability for settings where a more spatially accurate prediction is desired.

Quantitative Analysis Quantitative error analysis allows

users to compare the performance of different anomaly detection algorithms in a quantitative and objective manner. We provide the following metrics (cf. Fig. 1):

- Error-type distribution: How many of the errors made by the algorithms are of the types described in Sect. 3.1?
- IoU histogram: What is the distribution of IoU values? This histogram is a fine-granular view of the localization and detection ability of the algorithm.
- Oracle performance as explained in Sect. 3.2.

3.4 IMPLEMENTATION AND USAGE

TIADE implements a simple API for adding algorithms, their results, and starting the GUI or printing the results to standard output respectively. Implementation of TIADE was done with the simplicity of the API as a key principle. The overall tool was built as a single page web application using the StreamLit framework.

4 EVALUATION

In the following, we show-case the abilities of TIADE. An extensive comparison of different anomaly detection algorithms on challenging time-series datasets is left for future work and not in the scope of the paper.

```

1 import tiade as td
2 from tiade.data import data_loader
3
4 annotated_data = data_loader.load_cs_dataset(DATASET)()
5
6 evaluator = td.get_evaluator(annotated_data)
7 evaluator.add(method_identifier="mdi",
8               detections=predictions_of_mdi_algorithm)
9 evaluator.add(method_identifier="rrcf",
10              detections=predictions_of_rrcf_algorithm)
11
12 evaluator.start_gui()

```

Fig. 4. The API of TIADE: The "Evaluator" can be used to calculate and display both the comprehensive output for the webinterface, as well as a stripped down cli version

4.1 DATASET USED FOR SHOW-CASING TIADE

To evaluate TIADE, we used the The EEG Eye Artifact Dataset of [8], which contains multivariate EEG data of 45 sessions with 39 participants. On average, a trial holds 80 data carrying channels. The anomalies to detect are blinks, horizontal-, as well as vertical eye movement. Whereas the dataset is rather simple, it allows us to demonstrate the abilities of TIADE without focusing on a specific application and its challenges.

4.2 METHODS AND SETTINGS

Since in this paper, we show the general ability of TIADE for benchmarking instead of providing a comprehensive comparison of anomaly detection algorithms, we focused on detection results by Maximally Divergent Regions (MDR) algorithm presented in [3]. For the test, we used the Gaussian model and the unbiased KL divergence, without any preprocessing. In our evaluation, we set the IoU threshold to 0.5 and the score threshold to 0.0 in the settings panel of TIADE.

4.3 EVALUATION

For our evaluation, we focus on the quantitative results provided by TIADE, since some examples of the time series viewer are already shown in Fig. 1 and Fig. 4. Fig. 5 shows the error-type distribution as well as the mAP oracle. As can be seen in the right plot, fixing the missed GT issue would lead to a drastic improvement of the performance, although this is not directly visible in the error-type distribution plot. Furthermore, improving localization accuracy would be also beneficial, since there are a lot of near detections, which are also visible in the IoU histogram in Fig. 6. These insights would be not available by standard performance metrics, such as the recall-precision curve for a single IoU threshold.

In our evaluation, the provided metrics therefore proved useful in assessing the algorithm's strengths and weaknesses. The tool was also able to highlight the limitations of the algorithm's applicability given a diverse set of defect types.

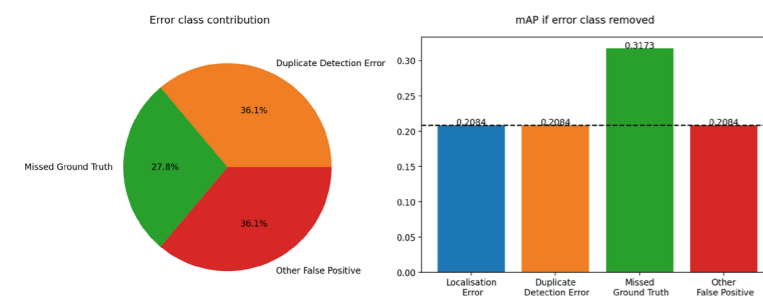


Fig. 5. Results for the EEG dataset and the blinks defect type: (Left) The error contribution diagram allows the user to compare the relative frequency of false detections. The improvement in mAP (right) shows that while missing ground truths is only responsible for 27.8% of all errors, increasing the detection threshold might be advisable - depending on the application.

5 CONCLUSION AND FUTURE WORK

In conclusion, the development of TIADE (Toolbox for Identifying Anomaly Detection Errors) offers an unprecedented advancement in the area of timeseries anomaly detection. This project goes beyond traditional evaluation metrics by allowing an in-depth, multi-factor analysis of anomaly detection algorithm performance. This intricate level of evaluation reveals significant insights into the strengths and weaknesses of algorithms, providing valuable feedback to refine the algorithms further. Our work has demonstrated that TIADE enables the differentiation of several error types. This functionality offers a superior analysis compared to single performance metrics. Through the utilization of TIADE, we can perform a more comprehensive evaluation, resulting in an enhanced understanding of the overall performance and the idiosyncrasies of various anomaly detection algorithms.

The application of TIADE to real-world data has revealed its practical effectiveness. By opening the project source code to the community on GitHub, we anticipate that researchers and practitioners alike can extend the utility of TIADE and continue to improve it. Looking towards the future, several opportunities exist for the expansion of TIADE. Firstly, although TIADE currently accommodates multi-variate timeseries data, there is scope to further enhance its

handling of diverse data types, such as heterogeneous time-series data with categorical variables, missing values or asynchronous acquisition times. Secondly, the graphical user interface could be enriched to provide more interactive and user-friendly visualizations, thereby making TIADE more accessible to users from various backgrounds. In conclusion, TIADE serves as a robust, detailed, and practical tool for time-series anomaly detection algorithm analysis. Its

development paves the way for future research in understanding and improving the performance of these algorithms.

Fig. 6. The IoU histogram, as taken from the results of evaluating the EEG dataset. This histogram allows the user to compare algorithms localization abilities. It is further indicative of how many predictions would classify as LocalizationError, depending on the chosen IoU threshold.

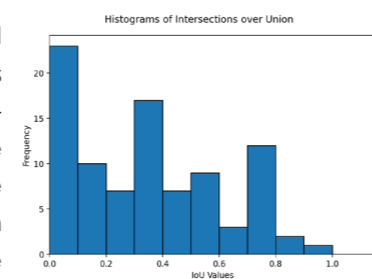


Fig. 6.

Acknowledgements: This work was supported by the KI-Werkstatt at the University of Applied Sciences Berlin (funded by German Federal Ministry of Education and Research, Forschung an Fachhochschulen, 13FH028K11) as well as the project TAHAI (funded by IFAF Berlin). We also thank Ricardo Knauer for fruitful discussions.

REFERENCES

- Aminikhanghahi, S., Cook, D.J.: A survey of methods for time series change point detection. *Knowledge and information systems* 51(2), 339–367 (2017)
- Barz, B., Garcia, Y.G., Rodner, E., Denzler, J.: Maximally divergent intervals for extreme weather event detection. In: *OCEANS 2017-Aberdeen*. pp. 1–9. IEEE (2017)
- Barz, B., Rodner, E., Garcia, Y.G., Denzler, J.: Detecting regions of maximal divergence for spatio-temporal anomaly detection. *IEEE transactions on pattern analysis and machine intelligence* 41(5), 1088–1101 (2018)
- Bolya, D., Foley, S., Hays, J., Hoffman, J.: Tide: A general toolbox for identifying object detection errors. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. pp. 558–573. Springer (2020)
- Hoiem, D., Chodpathumwan, Y., Dai, Q.: Diagnosing error in object detectors. In: *European conference on computer vision*. pp. 340–353. Springer (2012)
- Jia, W., Yang, L., Jia, Z., Zhao, W., Zhou, Y., Song, Q.: Tive: A toolbox for identifying video instance segmentation errors. *Neurocomputing* 545, 126321 (2023)
- Kim, J., Scott, C.D.: Robust kernel density estimation. *The Journal of Machine Learning Research* 13(1), 2529–2565 (2012)
- Kobler, R., Sburlea, A., Dias, C., Schwarz, A., Mondini, V., Müller-Putz, G.: EEG eye artifact dataset (Aug 2022). <https://doi.org/10.17605/OSF.IO/2QGRD>, <https://osf.io/2qgrd/>, publisher: Open Science Framework
- Ma, X., Zhang, Y., Xu, D., Zhou, D., Yi, S., Li, H., Ouyang, W.: Delving into localization errors for monocular 3d object detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 4721–4730 (2021)
- MacGregor, J.F., Kourti, T.: Statistical process control of multivariate processes. *Control engineering practice* 3(3), 403–414 (1995)
- Rewicki, F., Denzler, J., Niebling, J.: Is it worth it? comparing six deep and classical methods for unsupervised anomaly detection in time series. *Applied Sciences* 13(3), 1778 (2023)
- Tartakovsky, A.G., Rozovskii, B.L., Blazek, R.B., Kim, H.: A novel approach to detection of intrusions in computer networks via adaptive sequential and batchsequential change-point detection methods. *IEEE transactions on signal processing* 54(9), 3372–3382 (2006)
- Wang, A., Liu, A., Zhang, R., Kleiman, A., Kim, L., Zhao, D., Shirai, I., Narayanan, A., Russakovsky, O.: Revise: A tool for measuring and mitigating bias in visual datasets. *International Journal of Computer Vision* 130(7), 1790–1810 (2022)

Substitution of FEM simulations aided by AI

Robin Roj¹, Sebastian Chmielewski¹, Christoph Böttinger¹, Jens Kneifel¹, Ralf Theiß¹, and Peter Dültgen¹

¹ Forschungsgemeinschaft Werkzeuge und Werkstoffe e. V., Papenberger Str. 49, 42859 Remscheid

Abstract. In this work, an approach that employs artificial intelligence to support or replace certain finite element simulations is proposed, resulting in significant time and energy savings. An artificial intelligence model, based on autoencoders, is trained to predict simulation results for various loading scenarios. Using two two-dimensional examples of a clamped and loaded cantilever and a structural framework, it is demonstrated that generated images of displacement and von Mises stresses are consistent with the original simulation results. While the current implementation does not significantly reduce computation time, it represents a promising step towards integrating artificial intelligence into engineering software to support parameter finding during component design and reduce computation efforts. This work aims to enhance the efficiency and accuracy of product development processes, leading to potential energy and cost savings in industries that extensively use simulations. These require a lot of power and are performed iteratively due to constant changes to a technical component during the development phase. The approach aims to reduce the use of these computationally intensive simulations, or even substitute them, and accelerate the duration of iterative loops in the development process. Thus, the overall duration can be minimized, and resources can be conserved. For new geometries, e.g., small changes to a model, the neural networks are to substitute the simulation and make a prediction about the loads and deformations.

Keywords: Artificial Intelligence, Autoencoder, Computer Aided Design, Finite Element Method, Simulation.

1 INTRODUCTION

Typical approaches to increase efficiency or save energy mostly refer to the optimization of hardware and software, but not to the question of necessity. Due to the computational intensity of simulations with the finite element method (FEM), the fundamental question is whether they are necessary for every change in boundary conditions, such as geometries, grid structures, or loads. Energy saving potentials arise from the approach to support or even substitute those calculations by artificial intelligence (AI). The vision of this work is that after the training process, the neural networks can be applied to a new loading problem and are expected to provide a sufficient approximation of the simulation result in or with a fraction of the time and energy required by a conventional FEM simulation. Thus, the potential for significant energy and CO₂ savings arises in product development processes where FEM is applied, and it is intended to reduce the effort in generating FEM results or to substitute FEM calculations aided by AI.

The wide range of FEM simulations ensures that a substitution of time-consuming and computationally intensive FEM calculations by an AI would find a broad applicability in industry, research, and development in order to save energy at numerous companies and research institutions. E.g., a tool manufacturer can use the faster AI predictions within the same time it would take for an FEM calculation to validate more variations of the tools for strength, thus saving not only time and power, but also costs. Once the manufacturer has tapped into a suitable geometry based on the AI predictions, a final FEM simulation can be performed before the part is produced.

In particular, adapted geometries, such as variations of components, can be developed, designed, and validated in a significantly shorter time in the future using this method. For example, FEM simulations of customer-specific complex components currently running in parallel at tool manufacturers can still take several days per variant. Due to incremental changes in the development process of the component geometry, several FEM simulations have to be performed before the parts can be produced. Especially in such scenarios with different but similar variants, the use of AI enables potential savings.

2 STATE OF THE ART

The combination of AI with FEM has not yet been able to establish itself in professional proprietary CAD solutions (computer aided design), but has already produced promising approaches in research; especially for applications that require long computing times. Data mining methods are already being used, e.g., to simulate complex forging processes on the computer. The components are divided into a finite number of subareas (elements) and the software calculates displacements taking into account material strength, geometry of the die, and blank as well as many other factors [1].

A paper by Kononenko and Kononenko explores the possibility of augmenting FEM studies with machine learning to model physical systems [2]. FEM analyses are often used to solve differential equations with initial and boundary conditions, but solving large problems can be time consuming, taking hours to days. In addition, even minor adjustments to input parameters may require simulations to be created from scratch. The authors suggest that machine learning algorithms, if properly selected and trained, can improve upon conventional techniques. The paper discusses the benefits of machine learning in combination with FEM and gives some basic examples.

A paper by Zhou et al. focuses on learning latent representations of registered meshes for various 3D tasks [3]. The authors propose a non-template, fully convolutional mesh autoencoder that can process arbitrary registered mesh data. Conventional neuronal mesh autoencoders outperform conventional methods but have difficulty capturing fine-grained deformations. The model achieves higher reconstruction accuracy compared to modern methods, and its fully localized latent codes enable better interpolation capability than many conventional 3D mesh generation models.

In order to accelerate the time-consuming process of FEM simulation and to use resources more efficiently, scientists are working on a promising approach: the use of a self-learning AI-based algorithm [4]. Experts in solid forming are working together in an interdisciplinary way with specialists in AI to program an algorithm that can predict the results of FEM simulations. This algorithm is intended to recognize patterns from analyzed data sets and apply them to similar forming processes.

The goal was to optimize the algorithm so that it could predict the forming result in less than 60 s with a prediction accuracy of at least 95 %. This would significantly reduce the simulation effort and also keep detailed FEM simulation to a minimum. This could save engineers time and resources. Until now, engineers have designed their forgings in CAD programs and imported them into an FEM program for simulation. But with the help of the developed algorithm, forming prediction directly from the CAD program could be possible in the future.

Another advantage of the algorithm is its ability to automatically analyze complete test series, for example to test the forming of blanks with different lengths and diameters at different forming forces and temperatures. This would allow engineers to focus on those parameter combinations that look most promising in the prediction, which would significantly speed up the design of forming processes.

Thus, the integration of AI (autoencoders) into FEM software opens promising perspectives to increase efficiency, accuracy and competitiveness in the modeling and simulation of two-dimensional objects [5,6]. Future advances and research in this area will help to further develop this technology and expand its application areas in industry. Although some challenges, such as integrating AI into existing FEM software, ensuring data quality for training AI models, and ensuring transparency and explainability of predictions, still need to be overcome.

3 MATERIALS AND METHODS

For the generation and preparation of any input data to be fed into the AI in the next step, the software FreeCAD (version 0.19) was used [7]. It was assumed from the beginning that large amounts of data were necessary for the learning phase of the AI. The program is an open-source solution programmed with Python, which can be extended or automated by the community. For the automated creation of large amounts of data, the

so-called macro function was used. With this, design steps that are normally fed in by the user's manual input can be automated and thus massively accelerated. As described in the following two sections, it was decided to start with the simple example of a classic (two-dimensional) bending beam. Only in the second step was a more complex structure considered. For both variants, the design and the FEM calculation were automated using a macro function, so that the results were available in the form of text and CAD files.

These are shown in excerpts based on the cantilever in Fig. 1. In the first subframe, the three-dimensional geometry information of the meshed component is listed on the basis of x-, y-, and z-coordinates. The second subframe shows the displacement of each coordinate caused by the load. The third and fourth subframes list the stress and strain tensors. Other information not shown in the figure is given by the boundary conditions, such as the load application and restraint points. Thus, all information for the representation of the initial geometry, the boundary conditions as well as the FEM results are available in pure text form. In contrast to proprietary CAD systems, FreeCAD offers the advantage of being able to access and process all information.

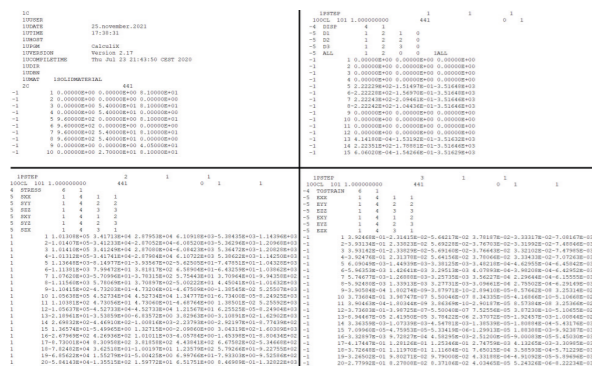


Fig. 1. Exemplarily text files generated with the FEM-module of FreeCAD.

3.1 CANTILEVER

As already mentioned, CAD models of simple bending beams were first used to test the entire workflow and generate initial results. Fig. 2 shows excerpts from the macro used for this purpose. First, a new CAD file was created using source code. In this, the „Box“ command was applied to achieve the basic shape of a bending beam. The elongation was set to generate a random value between 10 mm and 40 mm for the height, and a random value between 400 mm and 640 mm for the length. The width was defined with a fixed integer value of 100 mm.

The other entries, which can be seen in the source code of the macro, were used for the creation of the constraints, the application of a fixed force of 10,000 N, the definition of the material, the start of the FEM calculation, as well as the final saving of the file. As seen in the loop parameters, a test set of 10,000 files was created, which served for the training phase of the AI. The automated generation process took about 4 h.

```
for i in range(1,10000,1):
# new document
doc = App.newDocument("Beam")
outputpath = "C:\\FreeCAD-Automation\\Output"

# box as beam
box_obj = doc.addObject('Part::Box', 'Box')
box_obj.Height = random.randint(10,40)
box_obj.Width = int(100)
box_obj.Length = random.randint(400,640)
os.mkdir(outputpath+'\\'+str(i))
doc.saveAs(outputpath+'\\'+str(i)+'\\'+str(i)+'.FCStd')

# analysis
analysis_object = ObjectsFem.makeAnalysis(doc, "Analysis")

# material
material_object = ObjectsFem.makeMaterialSolid(doc, "SolidMaterial")
mat = material_object.Material
mat['Name'] = "Steel-Generich"
mat['YoungsModulus'] = "210000 MPa"
mat['PoissonRatio'] = "0.30"
mat['Density'] = "7900 kg/m^3"
material_object.Material = mat
analysis_object.addObject(material_object)

# fixed_constraint
fixed_con = ObjectsFem.makeConstraintFixed(doc, "FemConstraint-Fixed")
fixed_con.References = [(doc.Box, "Face1")]
analysis_object.addObject(fixed_con)

# force_constraint
force_con = ObjectsFem.makeConstraintForce(doc, "FemConstraint-Force")
force_con.References = [(doc.Box, "Face2")]
force_con.Force = int(10000)
force_con.Direction = (doc.Box, ["Edge5"])
force_con.Reversed = True
analysis_object.addObject(force_con)

# create mesh
femmesh_obj = ObjectsFem.makeMeshGmsh(doc, box_obj.Name + "_Mesh")
femmesh_obj.Part = doc.Box
gmsh_mesh = gmsh(femmesh_obj)
analysis_object.addObject(femmesh_obj)
doc.recompute()

# activating analysis
FemGui.setActiveAnalysis(doc.Analysis)
fea = cxxtools.FemToolsCxx()
fea.update_objects()
fea.setup_working_dir()
fea.setup_cxx()
message = fea.check_prerequisites()

# save
doc.saveAs(outputpath+'\\'+str(i)+'\\'+str(i)+'.FCStd')
App.closeDocument("Beam")
```

Fig. 2. Extract of macro source code for the generation of 10,000 cantilevers and FEM results.

Fig. 3 shows two examples of the results visualized in FreeCAD. It can be clearly seen that the width is identical at 100 mm, but the height and length differ greatly. At the bottom, the FEM results are displayed in the form of displacement. Since in the CAD environment, regardless of the load, each file is considered individually and thus the entire color spectrum is always filled, the coloring is intense in both examples. However, in order to represent the bending, this was optically intensified in the case of the thin beam in subframe 4, so that the displacement from the original position becomes clear.

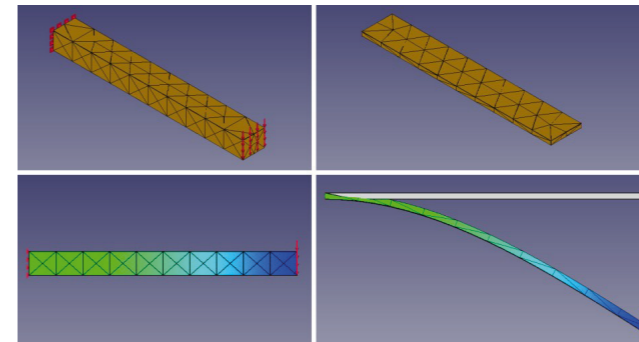


Fig. 3. Two examples of cantilevers with equal constraints and different geometry.

While the representations in Fig. 3 are only the visualization of the CAD models and the FEM results, the corresponding text files (cf. Fig. 1) were further processed using Python. The numerical values of the coordinates were extracted and transformed into simple images with black background. These images are shown by two examples in Fig. 4. As it can be seen in the bottom row, the images were created in such a way that the entire spectrum was not used in each file. On the contrary, the global maximum displacement was first determined from the entire test set of 10,000 files and the coloring of all images was oriented accordingly. This led to the effect that only elongated, thin bars with a strong deformation were colored intensely. The example with the thicker cantilever resulted in a perfect yellow coloration along the entire length.



Fig. 4. Two examples of monochrome and colored input images of the cantilevers.

The spectrum was adjusted so that the red value was varied between 0 and 255, while the green value was fixed at 255 and the blue value at 0. This resulted in a green-yellow gradient that was easy to recognize, with a strong green discoloration at the ends of the cantilevers with the greatest deformation. thus, each CAD file was the base for two images with a width of 640 pixels (corresponding to the maximum length of 640 mm), which were created in both monochrome and color. Both variants were necessary for the later input into the AI, since it was intended for the system to learn to colorize black-and-white images correctly.

3.2 STRUCTURAL FRAMEWORK

To prepare the investigations with a more complex example, a structural framework with a clamping on the lower left side and an applied force on the upper side was prepared. As with the bending beam, a macro was used to automatically generate 10,000 example files. However, the macro worked a little differently. First, the construction of one example was generated manually. The macro source code was then used to randomize various dimensions so that the shape of the component remained identical, but the distances between the struts, e.g., changed slightly. Fig. 5 shows only one example of the structure.

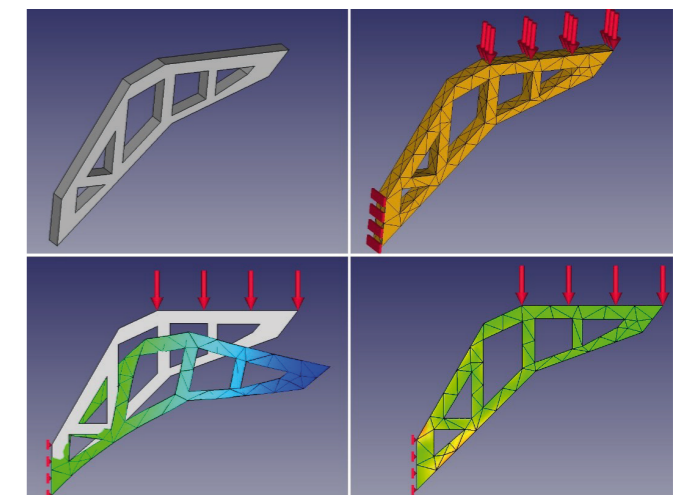


Fig. 5. One example of a structural framework showing the CAD model, the constraints as well as the FEM results in displacement and von Mises stresses.

In the first subframe, the original CAD model is shown. In subframe 2, the clamping point, the loading and the meshing were added. Subframe 3 shows the displacement with a downward deformation. In subframe 4, the von Mises stresses can be seen, with which the heavily loaded areas were colored red. Here, it is noticeable that the observer can predict the deformations (of bending beams and supporting structure) with ease; however, this is more difficult with the von Mises stresses. Also, from an optical point of view, no distinction is made here between compression and expansion. Similar to the bending beams, a corresponding number of input images of all 10,000 structures were prepared for the AI. In Fig. 6, the monochrome image is shown in subframe 1, the side view of the displacement in subframe 2, and the von Mises stresses in subframe 3. It becomes clear that the different intensities of coloration coincide with those of the CAD environment.



Fig. 6. Three images of the structural framework; the monochrome version, the displacement, and the von Mises stresses.

3.3 AUTOENCODER

As already mentioned in the state of the art, an AI tool based on autoencoders was predestinated to predict the coloration of the images. For this, a base code from Kaggle was selected [8]. After small adjustments in the code for the data preparation, as described in 3.1, the parameter for the encoder layers (down-sampling), the latent space or bottleneck, and the decoder layers (up-sampling) were set.

One layer setup was tested, and the input was inserted in the down-sampling layers (d1 to d5) to reduce the spatial dimensions and increase the depth of the input image, while the up-sampling layers (u1 to u5) aim to reconstruct the original input image by gradually restoring its spatial dimensions. The concatenation of feature maps from the down-sampling path to the up-sampling path at each level helps in preserving and combining the learned features. The final output was obtained through a convolutional layer with a kernel size of (3,3) and 3 output channels.

During the training phase for the cantilever and for the structural framework, both image sets were fed to the system. For this a training and testing split of 80 to 20 was implemented. After the training phase, the system was able to color monochrome image accordingly.

4 RESULTS

For the cantilevers, the system was expected to predict an only yellow coloration for beams with a comparatively large height. For long and thin cantilevers, a smooth color gradient from yellow to green would have been perfect. The results shown in Fig. 7 for both examples from Fig. 3 and Fig. 4 prove that the coloration for thick beams is working as expected with a correct yellow value along the whole length. In contrast to that, at first sight the thin beam is suddenly colored green at the end. Nevertheless, it is recognizable that slight green values are already appear before the sudden step.



Fig. 7. AI colored images of both example cantilevers.

In order to measure the results with a high accuracy, the original images have been compared with the predicted ones. As mentioned before, the green values were fixed to 255 and the blue values to 0. Thus, only the red values were varied, leading to the yellow appearance on the left. Table 1 shows the measured results of the red values along the whole cantilever. With a complete length of 628 pixels and mm, respectively, ten points have been considered in a distance of 70 units along the neutral fiber. On the right-hand side, the according red values are indicated in direct comparison of the original and the predicted image. Here, the sudden green coloration on the right tip can be recognized with the values of point 10. Nevertheless, the first five points can be evaluated as satisfactory.

Point	X	Red [original/predicted (deviation)]
1	1	255 / 254 (0.39 %)
2	70	254 / 235 (7.48 %)
3	140	248 / 235 (5.24 %)
4	210	238 / 234 (1.68 %)
5	280	232 / 233 (0.43 %)
6	350	212 / 233 (9.91 %)
7	420	196 / 234 (19.39 %)
8	490	179 / 235 (31.28 %)
9	560	161 / 235 (45.96 %)
10	628	149 / 184 (23.49 %)

In accordance with the cantilever, the approach for the structural framework was similar. It has been decided that the von Mises stresses were of greater interest since the displacement was trivial. Fig. 8 shows for direct comparison on the left-hand side the original image from Fig. 6 again. Here, ten measurement points have been defined at the areas of interest. On the right-hand side, the AI generated image of the structural framework is depicted, including the measurement points. At first sight, the coloration seems to be as intended.

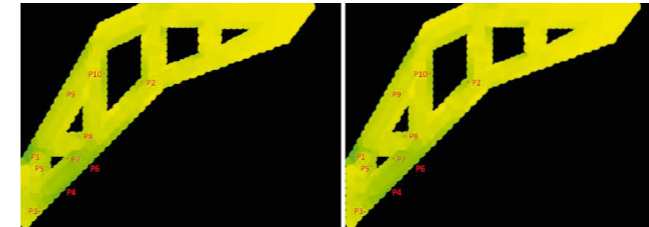


Fig. 8. Original and AI predicted image of the structural framework.

Table 2 indicates the exact position of the ten measurement points as x- and y-coordinated. Here, the point of origin is in the upper left corner. Similar to Table 1, the right column indicates the original and AI generated red values of all points. Only those have been chosen since the green and blue values were not relevant. In accordance with the optical impression of both images, the values show only small deviation, leading to the assumption that the coloration of further images also generates adequate results.

Point	X	y	Red [original/predicted (deviation)]
1	10	240	47 / 63 (34.04 %)
2	190	125	179 / 180 (0.59 %)
3	30	325	192 / 198 (3.13 %)
4	67	290	98 / 116 (18.37 %)
5	40	255	168 / 172 (2.38 %)
6	100	255	134 / 142 (5.97 %)
7	74	243	159 / 166 (4.40 %)
8	93	208	189 / 192 (1.59 %)
9	65	143	192 / 198 (3.13 %)
10	132	112	173 / 178 (2.89 %)

Fig. 4. Two examples of monochrome and colored input images of the cantilevers.

5 DISCUSSION AND CONCLUSIONS

In summary, the AI-generated colorization of the images produced higher quality results in the case of the structural framework and the von Mises stresses, respectively. In the case of the cantilever no smooth color gradient but a sudden green coloration at the end of the beam was achieved. In contrast to that, in the direct comparison of the original image of the structural framework with the predicted version, no difference could be optically recognized. Only the measured values showed a slight deviation, which, however, could be regarded as uncritical and satisfactory in the overall result. To recognize the deviation, the technique of analyzing the RGB-values of certain pixels has been applied. Since it was intended to

vary only the red value, it was possible to break down the quality measurement to only one significant indicator. Also, the percentages displayed in both tables, prove the variance to a larger or smaller value.

The reason for the comparatively low quality of the predicted images of the cantilever can only be assumed to be in connection with the smooth color change from one end to the other in the original image. This is not apparent in the case of the structural framework.

As described in the materials and methods section, the total amount of data was 10,000 CAD and image files, respectively. The three examples shown (thick bending beam, thin bending beam, and supporting structure) were selected in extracts to illustrate the overall results. It can be stated that they are representative for the prediction of all monochrome images.

Furthermore, it should be noted that the intention of the entire work was to use AI to reduce the effort in generating FEM results or to substitute FEM calculations. In the two examples of the cantilevers and the structural framework, an acceleration of the total time, a reduction of the computational effort, or a simplification for the user did not actually take place yet. With the value of 4 h mentioned before, for the FEM calculation of 10,000 CAD models of the cantilevers (including automated generation of the geometry), this results in a duration of 1.44 s per file. After completion of the training phase of the AI and feeding of the monochrome images into the pre-trained neural network, a duration of approximately 0.5 s per file for the generation of a prediction image was measured. With both values, it becomes clear that corresponding times do not play a major role for the everyday work of an engineer and the presented results represent only a first step in the direction of linking FEM and AI. This results in the following points, which are planned for future work:

- Development of metrics and tools for analyzing large amounts of data (e.g., for computation time measurement and quality analysis of predicted images)
- Consideration of more complex shapes and geometries
- Consideration of use cases that are more reflective of everyday engineering work and are based on real technical components
- Transfer of the method to three-dimensional use cases (the two selected examples of the cantilever and the structural framework had the advantage that a two-dimensional lateral view contained all necessary information)

- Feedback of the results into the CAD environment (the predicted colorations can theoretically be converted into stress and displacement values and thus visualized in the CAD software)
- Development of a software tool that supports the engineer in his daily work (e.g., which stresses arise at which locations when a geometric parameter is changed, or material is added at a different location)

Finally, it must be noted that for safety reasons, a complex FEM calculation should not be replaced by AI algorithms. The intention is rather an integration of AI into the CAD software, in order to support, in particular, the parameter finding process during the construction of components automatically and to reduce the effort of FEM computations. Once the geometry, all parameters, and the dimensions of a component have been finally defined, an extensive FEM calculation must follow in order to simulate the real application and to ensure safety.

ACKNOWLEDGEMENT

The progres.nrw – RESEARCH project „AI-Sim - Energieeinsparpotenziale durch die Annäherung von FEM-Simulationen durch künstliche Intelligenz“ (005-2211-0083_0363) is supported by the Ministry of Economic Affairs, Industry, Climate Action and Energy of the State of North Rhine-Westphalia, Germany. The authors thank the Project Management Jülich | Forschungszentrum Jülich GmbH.

REFERENCES

1. Rasche, N. F., Langner, J., Stonis, M., Behrens, B.-A.: Predicting Forming Forces and Lack of Volume with Data Mining Methods for a Flange Forging Process. *International Journal of Modeling and Optimization* 7(6), 363–369 (2017)
2. Kononenko, O., Kononenko, I.: Machine Learning and Finite Element Method for Physical Systems Modeling. <https://arxiv.org/abs/1801.07337>, last accessed 2023/07/22
3. Zhou, Y., Wu, C., Li, Z., Cao, C., Ye, Y., Saragih, J., Li, H., Sheikh, Y.: Fully convolutional mesh autoencoder using efficient spatially varying kernels. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., Lin, H. (eds.) *International Conference on Neural Information Processing Systems, NIPS'20*, vol. 776, pp. 9251–9262. Curran Associates, Inc., New York (2020)
4. IPH Institut für integrierte Produktion Hannover, https://www.iph-hannover.de/de/for-schung/forschungsprojekte/?we_objectID=2316, last accessed 2023/07/21
5. Pyimagesearch, <https://pyimagesearch.com/2020/02/17/autoencoders-with-keras-tensor-flow-and-deep-learning/>, last accessed 2023/07/21
6. Tensorflow, <https://www.tensorflow.org/tutorials/generative/autoencoder>, last accessed 2023/07/21
7. FreeCAD, <https://www.freecad.org/index.php>, last accessed 2023/07/21
8. Kaggle, <https://www.kaggle.com/code/theblackmamba31/autoencoder-grayscale-to-color-image>, last accessed 2023/07/21

Classification of stone samples for concrete production using hyperspectral imaging and deep learning

Daniel Garten¹, Katharina Anding^{2,3}, Galina Polte², Elske Linß⁴, Raphael Gerlach⁵

¹ GFE - Gesellschaft für Fertigungstechnik und Entwicklung Schmalkalden e. V.

² Technische Universität Ilmenau, Fachgebiet Qualitätssicherung und Industrielle Bildverarbeitung

³ Now with Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau

⁴ MFPA - Materialforschungs- und -prüfanstalt an der Bauhaus-Universität Weimar

⁵ SQB - Steinbeis Qualitätssicherung Bildverarbeitung GmbH, Ilmenau

Abstract. An optical system for the detection of concrete-damaging components in aggregates will be presented. This is intended for use in the construction industry and is based on hyperspectral imaging and automated real-time hyperspectral image analysis. It uses machine learning methods and neural network architectures to detect relevant concrete-damaging components in the aggregates. Currently, optical analysis and sorting methods are not yet used on a large scale in aggregate processing. This is due in particular to the high complexity of the detection problem with a high phenotypic variability within the classes to be distinguished and high interclass variability. In the given case, a regular random analysis of representative samples is carried out for each unconsolidated rock deposit whose decomposition product is used for concrete production. The sample is separated and transported by an adapted object feed and object separation system, consisting of a vibrating conveyor for coarse separation and a subsequent conveyor belt for stable RGB and hyperspectral data acquisition. A high-resolution RGB color line scan camera (4096 pixels) and a hyperspectral camera with sensitivity in the 900 - 1700 nm wavelength band are used for synchronous data acquisition. Furthermore, deep learning methods as well as conventional methods such as support vector machine and random forest classifier are presented for the classification of hyperspectral data. In particular, the transformation of the data, e.g. by means of LDA (linear discriminant analysis) and PCA (principal component analysis), is discussed in order to make it usable for network architectures designed for RGB color images. Different approaches for combining the spectral data and the higher-resolution RGB data from the color line scan camera (high information content for texture analysis) are also discussed. As a result, a recognition accuracy of more than 90 % was achieved for the highly complex recognition problem.

1 INTRODUCTION

Natural aggregates, which are used as concrete aggregates, have to be analysed according to certain European standards, in Germany also according to relevant federal standards, for example [1] in Saxony-Anhalt. Samples from natural mineral deposits need to be taken at frequent intervals for the analysis of mineral components. Samples must be taken from natural mineral deposits at regular intervals for petrographic analysis of the mineral components. Deposits that exceed the specified limits for the presence of certain minerals are not suitable. In Germany and also worldwide, a large number of structural damages are caused by the use of constituents in natural aggregates that are harmful to concrete. In Germany, for example, about 400 km of motorways are affected by the alkali-silica reaction (ASR) [2]. A damage pattern of alkali-silica reaction that can be caused by the use of alkali-silica reactive aggregate as concrete aggregate is shown in the example of a bridge in Figure 1.

Another problem is that in addition to the alkali-sensitive aggregates, there are other aggregate components that negatively affect the performance properties, durability and also the visual aspect of the concrete surfaces in very different ways.

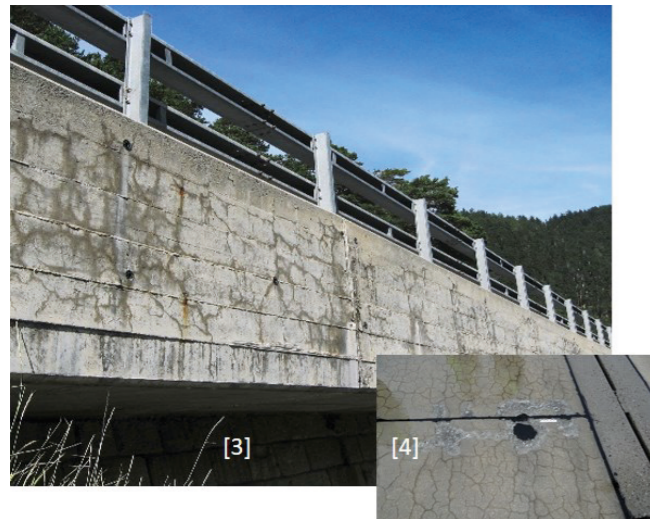


Fig. 1. Damage pattern of the alkali-silica reaction [3], [4].

The state of the art for a petrographic analysis of natural aggregates is still a manual examination by specially trained experts (such as mineralogists or geologists). This procedure is time-consuming, error-prone and depends on the individual experience of the inspector. Automatic routines can only be found for the analysis of few specific mineral components (for example Calcite and Dolomite) by using the NIR spectra [5]. But the given natural aggregates used as concrete additives are already present within individual classes as mixed mineral components with specific chemical and spectral properties (a stone particle often contains several mineral components in its structure). Therefore, determining the class of stone particles using NIR or SWIR characteristics is more difficult than separating pure minerals. When analysing the natural aggregates to be classified, the difficulty of the task thus lies in a large intra-class variability with a simultaneously low inter-class variability. In concrete terms, this means that the individual class shows strong variations in appearance and the differences in appearance between the classes are unfortunately often very small. This results in possible confusion between the classes. Examples of stone classes with very similar appearance are shown in the figure 2 with their colour image.

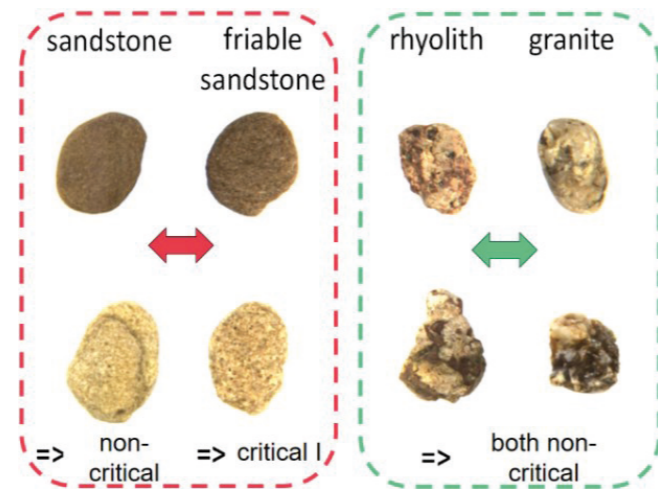


Fig. 2. Obviously very similar objects belonging to different gravel types.

Wotruba et al. wrote that the number of minerals, which can be automatically analysed in NIR, is limited because not all minerals showing diagnostic absorption features and very dark minerals can also cause problems by absorbing a high amount of the electro-magnetic spectra [5]. The image analysis of grinding surface patterns is another way for automatic classifying of natural aggregates [6]. This procedure is used for the inspection of grinding surface patterns of mineral plates. This procedure needs a costly mechanical processing of the mineral surface. This is not useful for the classification of natural aggregates as untreated stone material. This paper describes an automatic recognition routine for petrographic inspection by using image processing of hyperspectral imaging data and artificial intelligence algorithms in the form of pre-trained deep learning networks in the SWIR spectrum. In our own previous studies, we used image processing and machine learning based on colour images in the VIS spectrum [7], [8], [9].

The results so far have already been good, but especially the differentiation of very similar classes and their correct assignment to the superclasses „non-critical“, „critical I“ and „critical II“ could not yet be solved sufficiently well on the basis of the VIS colour image. For this reason, further investigations in SWIR were continued in the form of hyperspectral data analysis.

2 HARDWARE SETUP

The sample components are separated and transported by an adapted object feed and object separation system, consisting of a vibrating conveyor for coarse separation and a subsequent conveyor belt for stable RGB and hyperspectral data acquisition. A high-resolution RGB color line scan camera (4096 pixels) and a hyperspectral camera with sensitivity in the 900 - 1700 nm wavelength band are used for synchronous data acquisition. In order to prevent vibrations from being coupled into the camera systems, both the line scan and hyperspectral cameras are mounted on vibration-damped bearings. Two halogen light sources are used in the hyperspectral camera's field of view to achieve shadow-free illumination. Two LED bar lights are used for the color line scan camera. The temperature and color spectrum of both systems are monitored by sensors and their intensity can be adjusted. The control is carried out via a DLL (dynamic link library) from the control program of the test system.

The camera systems were adjusted in terms of line length and synchronized to a recording frequency of 500 Hz. The complete setup is shown in figure 3.

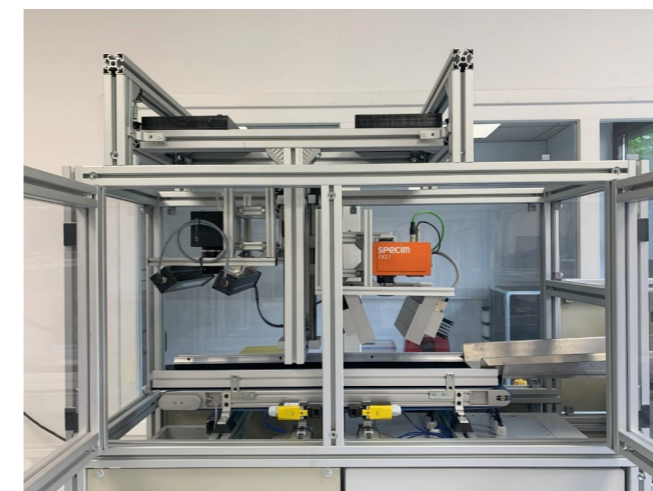


Fig. 3. Hardware setup for sample component separation, transportation and image acquisition

3 DEEP LEARNING ON HYPERSPECTRAL DATA

Known network architectures of the CNN (Convolutional Neural Network) expect single or three-channel images. This poses a problem for the processing of multi-channel, in our case 255-channel images. One possible solution is dimension reduction using Principal Component Analysis (PCA), Independent Component Analysis (ICA) or Linear Discriminant Analysis (LDA). So, new features

are created from a combination of the original features space. However, only the most important information in the new features is taken into account in order to obtain the most accurate representation of the data. In order to get an overview of the distribution of the data and the formation of clusters in high-dimensional space, point spectra of a data set with all 21 classes (which we later summarize into 3 relevant super-groups to reduce complexity) were subjected to LDA. Figure 4 shows the results of the dimension reduction to three dimensions using LDA. It is clearly recognizable that clusters are formed and that it is therefore possible to recognize the individual sample components on the basis of the point spectra. Algorithms for baseline correction, normalization, outlier analysis and elimination as well as data reduction (intelligent channel selection, feature space transformation) were developed and examined and validated on the basis of the given HSI data sets.) In addition, pre-processing steps were analyzed for point spectra of individual pixels of selected objects from the HSI sensor signal. The whole class structure with 21 sub-classes is shown in figure 5.

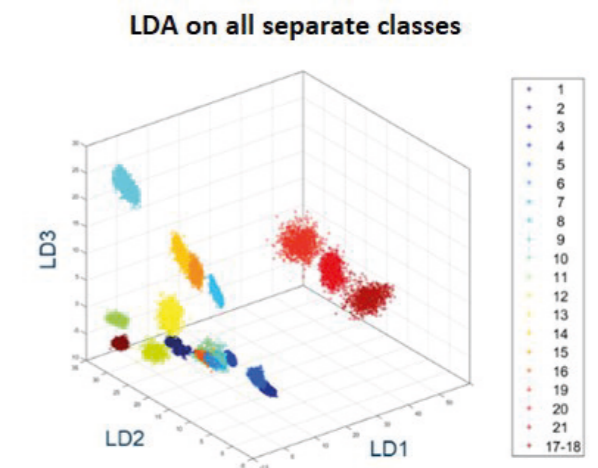


Fig. 4. LDA on all separate classes

Category	Aggregates	Name of class	
1	Quarz	non-critical aggregates	
2	Siliceous slate (black, grey)		
3	Quartzite		
5	Other Palaeozoic sediments (quartzitic, clay, phyllitic schists)		
6	Sandstone (dense) except group 16		
7	Limestone (marlstone), native except group 15**		
8	Limestone (dolomite), Nordic except group 15**		
10	Crystalline (granite, gneiss), Nordic		
11	Flint (dense), all varieties except group 12*		
4	Greywacke		critical aggregates 1 (Water-absorbent, swellable, inorganic, alkali-reactive)
9	Rhyolite, andesite, (porphyries, porphyrites), basic volcanic rocks		
12	Chalk-crusted and porous flint (flint)*		
13	Siliceous limestones, siliceous chalk, opaline sandstone		
14	Chalk / Chalk Limes		
15	Light and porous limestone and Sedimentary rocks with loose grain bonding (e.g. mudstone / siltstone /		
16	Brown coal	critical aggregates 2	
17	Charcoal wood, xylitol		
21-1	Other - plastics, rubber, polystyrene		
21-2	Other - organic material (plant parts, soil, ...)		
19	Brown iron encrustations, turf iron ore		
21-3	Other - Goethite		
21-4	Other - Haematite		
21-5	Other - metals		
21-6	Other - building materials (mortar, concrete, bricks,...)		
20	Pyrite, marcasite		
21-7	Other - gypsum		

Fig. 5. Whole class structure with 21 sub-classes

For the classification of these dimension-reduced hyperspectral images, three pretrained deep-learning models, VGG16, ResNet50 and MobileNet, were used, which were specially developed for the classification of 3-channel image data. For building the partitions of the training, validation and test dataset, a dataset partitioning ratio of 70/15/15 was used. The best results are achieved with the image data generated on the basis of the supervised dimensionality reduction method LDA. This supervised dimensionality reduction method, which takes class-specific characteristics into account as prior knowledge, aims to increase the distance between classes and decrease the distance between data within each class. The resulting three channel images of sample objects after dimension reduction are shown in figure 6.

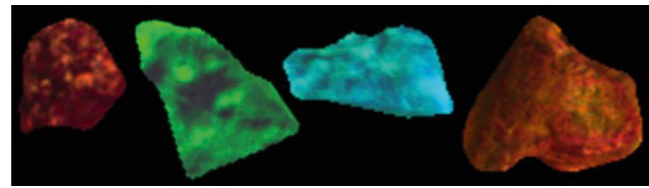


Fig. 6. The resulting three channel images of sample objects after dimension reduction

Fusion of the classification results obtained from the hyperspectral and RGB color line scan camera data is a complex process. Joint processing of the data in a single network is not expedient due to different fields of view and spatial resolutions.

The strategy here is to train two different classifiers. Preliminary investigations have shown that higher recognition accuracies can be achieved on the basis of hyperspectral data. Therefore, in the first stage, a classification is carried out using CNN on the hyperspectral data and if the confidence falls below a defined level, the object in question is searched for in the data stream of the color line scan camera and a further classification is carried out based on the RGB data. The result with the higher confidence is finally adopted as the classification decision.

4 COMPARISON OF DEEP LEARNING AND CONVENTIONAL MACHINE LEARNING METHODS ON THE BASIS OF HYPERSPECTRAL SENSOR DATA

In order to gain an insight into possible differences between the results of conventional feature-based learning methods and deep learning, different methods were compared on two datasets. The best results were achieved with a random forest classifier and a ResNet-50. For the random forest classifier, color- and texture-based features were calculated on the basis of three LDA-reduced channels (LDA - linear discriminant analysis). These are statistical color features such as mean and standard deviation in a specific channel as well as texture parameters according to Laws [10]. The results are shown in figure 7.

superclass	classified as			Sum	Recognition Rate	
	uncritical	critical I	critical II			
reals	uncritical	625	62	0	687	90,98%
	critical I	72	552	6	630	87,62%
	critical II	1	7	285	293	97,27%
				1610	90,81%	

Fig. 7. Best classification results on dataset 1 (random forest)

The results obtained with CNN using the LDA-reduced 3-channel image as input and an extended data set 2 are shown in figure 8.

superclass	classified as			Sum	Recognition Rate	
	uncritical	critical I	critical II			
reals	uncritical	1073	155	0	1228	87,38%
	critical I	155	1080	3	1238	87,24%
	critical II	0	6	471	477	98,74%
				2943	89,16%	

Fig. 8. Best classification results on dataset 2 (CNN)

The results show that the deep-learning-based method provided the best results on dataset 2, however slightly lower recognition accuracies than the random-forest classifier on dataset 1. An increase in recognition quality can be expected by adjusting the parameters of the network. The different size of the dataset should be also considered into account.

It is assumed that current investigations into cascaded classification using the different sensor data of RGB color camera sensor and HSI sensor will lead to a further increase in the recognition rate.

For the investigations in this paper the image processing library MVTec Halcon Version 20.55 [11] was used.

5 CONCLUSION

In this paper, a method for the automatic inspection of natural stone aggregates using hyperspectral imaging and AI algorithms in the form of deep learning networks could be demonstrated. With an overall recognition rate of around 90 %, the deep learning model known as ResNet50 in combination with LDA as a dimension reduction method proved to be particularly well suited. Through so-called fine-tuning, the overall recognition rate could be raised further. Overall, however, there is a need for optimization within the process chain for the classification of hyperspectral images, which concerns the image acquisition technique, image pre-processing and segmentation as well as dimension reduction and the learning rate. It is to be expected that the use of high-resolution texture information from the line scan camera image will lead to an increase in the recognition rate. This is the main focus of the current investigations.

6 ACKNOWLEDGEMENTS

The underlying research and development project „Automated real-time hyperspectral imaging and analysis for the detection of concrete-damaging constituents in aggregates - HyPetro“ is funded by the Federal Ministry of Education and Research (BMBF) as part of the funding program „Computer-Aided Photonics - Holistic system solutions from photonic processes and digital information processing“ and supported by the project management organization VDI Technologiezentrum GmbH. Special thanks go to the funding bodies that made it possible to carry out the research on the above mentioned topic in the first place through their financial support.

The responsibility for the research content lies with the authors.

REFERENCES

- ZTV-StB LBB LSA 09/10 – Additionally technical contract conditions and guidance for road construction for the business area State Office for Building Construction of Saxony-Anhalt, 2009.
- Ludwig, H.-M.: FUTURUM – Baustoff Straßenbeton. Forschungskolloquium „Bet-onstraßenbau“, 2014
- Schadensbild der Alkali-Kieselsäurereaktion, <https://www.uni-weimar.de/de/bauingenieurwesen/professuren/werkstoffe-des-bauens/struktur-arbeitsgruppen/alkali-kiesel-saeure-reaktion/>
- Schadensbild der Alkali-Kieselsäurereaktion, <https://www.baublatt.ch/baupraxis/aar-empa-und-epfl-der-betonkrankheit-auf-der-spur-25152>
- H. Wotruba, M. R. Robben, D. Balthasar: Near-Infrared Sensor-Based Sorting in the Minerals Industry, Conference in Minerals Engineering, Lulea, Schweden, 2009.
- L. Lepistö: Colour and Texture Based Classification of Rock Images Using Classifier Combinations, doctoral dissertation, Tampere University of Technology, 2006.
- K. Anding, D. Garten, G. Linß, G. Pieper, E. Linß: Classification of mineral construction raw materials from unconsolidated mineral deposits by means of image processing and machine learning (Klassifikation mineralischer Baurohstoffe aus Lockergesteinslagerstätten mittels Bildverarbeitung und maschinellem Lernen). 16th Colour-Workshop of the German Color Group, 2010.
- Linß, G., Anding, K., Garten, D., Göpfert, A., Reetz, E. and Rückwardt, M.: Auto-matic Petrographic Inspection by Using Image Processing and Machine Learning, XX IMEKO World Congress, Metrology for Green Growth, Busan, Republic of Korea, September 2012.
- Anding, K., Haar, L., Polte, G., Watz, J., Notni, G.: Comparison of the performance of innovative deep learning and classical methods of machine learning to solve industrial recognition tasks, In: IMEKO TC1 – TC2 International Symposium on Photonics and Education in Measurement Science 2019, 17.09.-19.09.2019.
- Laws, K. I.: Textured image segmentation. Image Processing Institute, University of Southern California. IPI Report 940, 1980, 186 p.
- Halcon 20.05, Comprehensive standard software for machine vision with an integrated development environment, <https://www.mvtec.com/de/produkte/halcon/>, 2022.



**25. Anwendungsbezogener Workshop zur
Erfassung, Modellierung, Verarbeitung und Auswertung von 3D-Daten**

Berlin
30.11./ 01.12.2023
Gesellschaft zur Förderung
angewandter Informatik e. V.

ELV.xD - a modular framework for instance level recognition within spare part logistics.

Marian Bookhahn¹, Kevin Fuge¹, Gabriel Thormann², Charleen Möhr², Kai Zetzmann², Holger Lürer², Mario Olschowski², Frank Neumann¹

¹HTW – Hochschule für Technik und Wirtschaft Berlin
Fachbereich 2: Ingenieurwissenschaften - Technik und Leben
Studiengang Ingenieurinformatik
Wilhelminenhofstr. 75a, 12459 Berlin
eMail: frank.neumann@htw-berlin.de

²PI Informatik GmbH, Hertzstraße 67, 13158 Berlin

Abstract. The research project ELV.xD aims at providing a modular frame-work for instance level recognition to spare parts logistics. In this use case, complex parts and spare parts have to be identified within a catalog consisting of several 10,000 up to 100,000 of components having close geometrical similarities. ELV.xD employs the RecoStation appliance equipped with cameras and depth sensors to determine the most similar component within the inventory. The project aims at providing a modular framework for instance level recognition by combining 2D image-based and supervised or neural network-based methods with 3D-based recognition methods. The present paper focusses on the 3D part of the framework and describes the different analysis and experiments steps conducted to find optimal solutions for the 3D pipeline. In detail, it demonstrates that the 3D reconstruction with Intel Realsense D415 depth sensors can be improved by replacing Realsense's on-board stereo matching with the HITNet method. On average the HITNet supported reconstruction pipeline is able to reduce the RMSE of an ICP registration between a reconstructed and a ground truth shape by over 25%. Finally, it introduces the instance retrieval algorithm and evaluates the performance of the overall 3D pipeline within ELV.xD. The results suggest, that the HITNet aided reconstruction method can significantly improve the overall recall performance.

1 INTRODUCTION

A major challenge in the context of geometry based industrial object recognition is the performant determination of concrete, complex parts from a data set comprising tens or even hundreds of thousands of parts, which may exist in representative part categories or have a high degree of similarity (e.g., valves, flanges, hydraulic fittings, etc.). Thus, the size of the search space makes classical ICP-based approaches impracticable for such a task. When using optical sensors to capture 2d and 3d data of these objects, additional challenges to consider are different viewpoints for every capture, different textures for geometrically identical parts as well as the acquisition speed.

In this paper we propose a highly performant system, capable of accurate 3D reconstruction and instance level recognition built from relatively inexpensive hardware components. Specifically, we demonstrate an optimized 3d reconstruction pipeline based on active stereo cameras like the Intel Realsense D415. In addition, we show how specifically adapted pointcloud based instance retrieval methods benefit from the optimized reconstruction pipeline by evaluating the recall performance of different stages of optimization.



Figure 1: RecoStation II

2 RELATED WORK

Stereo Matching

Binocular stereo matching has been researched for decades. With the use of handcrafted heuristics, classical approaches find local correspondences. Where local methods stop by simply searching every pixel, global methods try to make use of spatial context by incorporating a global energy function. In recent years, learning-based methods brought remarkable improvements to this area. At the beginning, most of these approaches focused on extracting patch-wise features with siamese networks. Later various works proposed

end-to-end architectures which improved the accuracy even further.

To mimic the filtering stages in classical stereo matching algorithms like SGM [1], the majority of recent approaches utilize a cost volume layer introduced by [2], to which a series of 3D convolutions is applied. Based on this work, several models use cascaded cost volumes [3] [4] [5] [6] [7] [8] to reduce expensive 3D operations. HITNet [9] uses a different approach to be more efficient. Inspired by classical stereo matching methods, it first extracts compact features to initialize multi-resolution disparity maps and refines the estimates in a 2D disparity propagation stage using slanted support windows and learnable features.

3D Reconstruction

3D reconstruction describes the process of capturing the shape and surface characteristics of physical components starting with the input data of different sensors (e.g. RGB cameras, ToF sensors, stereo cameras, structured light sensors, 3D scanners). The goal of the 3D reconstruction process is the generation of a 3D model (typically a mesh) with very high geometrical and visual similarity to the captured component. The various reconstruction methods differ in the input data used, the required processing time, and the achieved reconstruction quality. Photogrammetry focusses on the calculation of 3D data based on a collection of photographs of a component taken from different viewpoints. Based on photogrammetry, Apple recently added the Object Capture API [10] to specific iPhones and iPads. Both the capturing and the calculation process for a given object require a couple of minutes and should be considered as rather time consuming. Other approaches process depth images provided by ToF sensors, stereo cameras or structured light sensors. A typical example is the two-stage approach [11] adopted in the research project SparePartAssist, where the first stage fuses the various depth images into a Truncated Signed Distance Function (TSDF) representation. Subsequently, a second stage generates a triangle mesh based on the TSDF representation. For this purpose, dedicated versions of the marching cubes algorithms are applied [11]. By combining the two stages near real-time performance for 3D reconstruction may be achieved.

3D Instance Level Recognition

To recognize an object in a big collection of shapes, usually a similarity distance function is incorporated. The advances in autonomous driving have pushed the development of deep learning methods for point cloud based similarity search for place recognition in recent years.

As a result of this trend, PointNetVLAD [12] has emerged, combining the strengths of PointNet [13] and NetVLAD [14] to predict a VLAD descriptor for point clouds in an end-to-end fashion. PCAN [15] builds upon this approach by modifying Point-Net++ [16] to generate an attention map.

In contrast, LPD-Net [17] takes a different approach, still relying on feature aggregation through NetVLAD [14], it uses handcrafted local features and graph neural networks to extract local contextual information. Building upon the foundations of PointNetVLAD [12], SOE-Net [18] incorporates self-attention mechanisms to learn long-range context dependencies and seamlessly integrates local descriptor extraction and aggregation, allowing for one-stage training to produce a global descriptor.

Another method, MinkLoc3Dv2 [19], begins by quantizing the input point cloud into a sparse voxelized representation. It then adapts the Feature Pyramid Network to extract local features and employs a Generalized-Mean pooling layer to aggregate these local features into a global descriptor. This approach has proven to achieve state-of-the-art performance in the realm of place recognition tasks.

3 METHODS

The PI Informatik GmbH has developed an innovative system for multi-dimensional master data capture and the consolidation of historically grown, heterogeneous data repositories with the RecoStation. In pursuit of a simple application-specific adaptability, the system was built in a scalable manner based on established, market-standard, and cost-effective components. Relevant components include Intel Realsense technologies, standards such as Nvidia graphics cards, and Intel Core CPU technology, among others.

The process support for spare parts management is carried out through a high-performance one-step procedure. In this procedure, visual and alphanumeric data from various source systems are merged and evaluated using ML techniques.

While the previous focus in data analysis was oriented towards the 2D world, the present research work has extended this focus to the 3D world and incorporated significant innovative approaches to multidimensional, ML-based data processing. We aim at improving the quality and performance of recognition by combining 2D image-based and neural network-based methods with 3D-based recognition methods for instance-level recognition.

Taking this into account, the RecoStation utilizes 9 active stereo cameras (Intel Realsense D415) surrounding the acquisition space (Figure 2) in order to capture geometry and color information. As no moving parts are included, the recording speed is only limited by the sensor's frame rate, the number of frames used for temporal filtering and data transfer. Sensor and process related inaccuracies (e.g., temporal and spatial noise, stereo matching, calibration, ...) are affecting the 3D reconstruction.

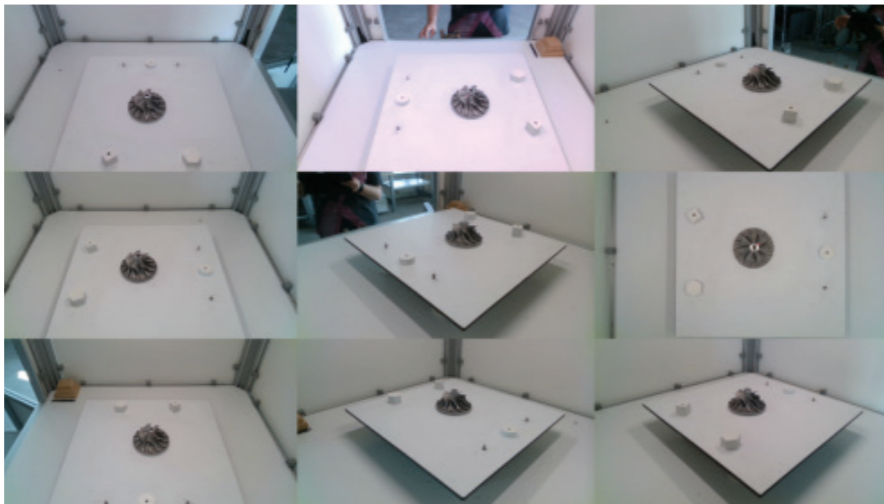


Figure 2: The views of RecoStation's 9 cameras.

Considering 3D recognition benefits from accurate 3D reconstruction, we analyzed how to improve the reconstruction process given the sensor setup is already fixed. One promising approach is to replace the sensor's onboard stereo matching process with a deep learning method, since in this area remarkable progress has been made in recent years.

RecoStation's search capability is designed to be highly performant, even with collections of up to 100,000 components. Thus, classical ICP-based approaches are too slow for such a large search space. The vast majority of modern search algorithms are incorporating a compressed description of the objects, which can be compared very efficiently.

In this way, several point cloud based deep learning methods have been proposed in recent years for the place recognition task in autonomous driving. Considering these methods are solving a problem very similar to our task, we investigate how to utilize these approaches.

The following sections will primarily highlight the previous results of the 3D-based methods.

THE CENTRAL AUTHORITY: THE MASTER

The 3D procedures are integrated into a system concept controlled by a central authority, known as the Master. The Master specifically manages the data capture systems, the data capture process for the parts, and oversees the spectrum of possible recognition methods for the specific recognition process. By integrating and parallelizing multiple recognition methods, the quality of results can be compared, and benchmarks can be derived. For this purpose, the results are presented in a comparable manner using a simple visualization concept (e.g., using charts, and more). To illustrate comparability and result assessment, a generic pipeline concept has been developed alongside the 3D pipeline, which serves as a basis for integrating additional AI methods. Currently, several optional AI methods are being examined, orchestrated, and established as benchmarks. Therefore, the range of procedures can be dynamically expanded as needed.

THE 3D APPROACH MODEL / 3D PIPELINE

Starting from the data set generated by the source sys-

tems for each individual object, point clouds are created, which are then passed to the specific AI process. Using this process, object-specific descriptors are generated, which can be compared with those of objects already existing in the inventory. Specifically, the following pipeline was implemented:

Step 01: Depth Estimation

The Intel Realsense D415 stereo camera offers a maximum depth image resolution of 1280 x 720 pixels, while its raw infrared images are captured at a resolution of 1920 x 1080 pixels. Since we are not dependent on Realsense built-in stereo matching, we can integrate the full resolution of the raw infrared images. To obtain depth information, we feed each stereo image pair to HITNet, which estimates the stereo disparities from which we calculate all depth values.

Step 02: Reconstruction

For 3D reconstruction, all depth images are fed into the TSDF pipeline of Open3D [20]. The use of three consecutive images taken from each sensor dramatically reduces the influence of temporal noise. As a quality measure, we calculate the Root Mean Square Error (RMSE) of an Iterative Closest Point (ICP) registration between a reconstructed and a ground truth shape in section 4.

Step 03: Descriptor Extraction

To obtain compressed descriptions that maintain consistency across changes in viewpoint for each reconstructed mesh, point clouds are sampled with a size of 4096 via Poisson Disk Sampling and then input into a specially trained version of PointNetVLAD [12]. The retrieved shape descriptors are stored in a database. To search for an object based on its shape, its descriptor is determined by the system and then compared to all database descriptors using cosine similarity (normalized dot product of two vectors).

The management of descriptors in the database is handled by the Master and provided contextually for the respective data capture process individually. We evaluate the performance of this stage by calculating the Recall for up to 15 nearest neighbors with:

$$Recall = \left(\frac{True\ Positives}{True\ Positives + False\ Negatives} \right) \quad (1)$$

4 RESULTS AND DISCUSSION

Throughout all of the following experiments the capturing related parameters (eg. exposure time, infrared projector power and lighting) has been kept constant. The objects were placed manually in an arbitrary position into the acquisition space and

3D Reconstruction

The qualitative comparison of the reconstruction results depicted in Figure 3 demonstrates clearly, how the reconstruction pipeline benefits from HIT-Net's depth estimates. With sharper edges and fewer noise, the reconstructed mesh quality improved notably.

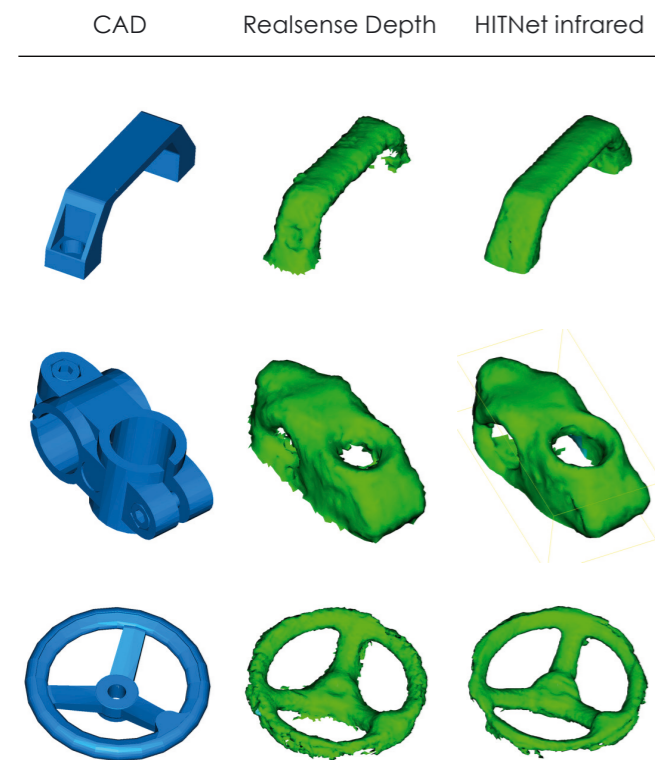


Figure 3: Comparison of reconstruction results using the genuine Realsense and the HITNet depth estimation.

To evaluate how the reconstruction quality improves by using HITNet [9], we calculate the RMSE for registering several reconstructed meshes to their reference CAD models using CloudCompare's [21] build-in ICP (fine registration). As shown in Figure 4, in all experiments the HITNet reconstruction pipeline achieves lower RMSE values. On average the HITNet supported reconstruction pipeline is able to reduce the RMSE by over 25%.

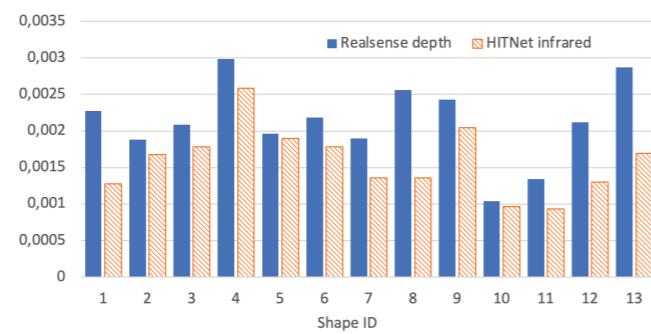


Figure 4: Comparison of the ICP registration RMSE for two different reconstruction pipelines. Blue: Open3D TSDf + Realsense D415 depth frames. Orange: Open3D TSDf + HITNet depth estimation on Realsense D415 infrared frames.

3D Instance Level Recognition

To evaluate the performance of our shape similarity search, we store the descriptors of 37 shapes in our database and take 3 additional captures of each shape with distinct poses. As our shape similarity search is point cloud based, we analyze 4 different approaches to reconstruct point clouds.

1. **Realsense:** Point cloud aggregation from Realsense depth images + Farthest Point Sampling
2. **HITNET:** Point cloud aggregation from HITNet depth estimates + Farthest Point Sampling
3. **Realsense + TSDf:** TSDf reconstruction with Realsense depth images + Poisson Disk Sampling
4. **HITNET + TSDf:** TSDf reconstruction with Realsense depth images + Poisson Disk Sampling

For these 111 captures we calculate the recall for up to 15 retrievals. Figure 5 shows how our shape similarity search clearly benefits from an improved reconstruction quality using HITNet.

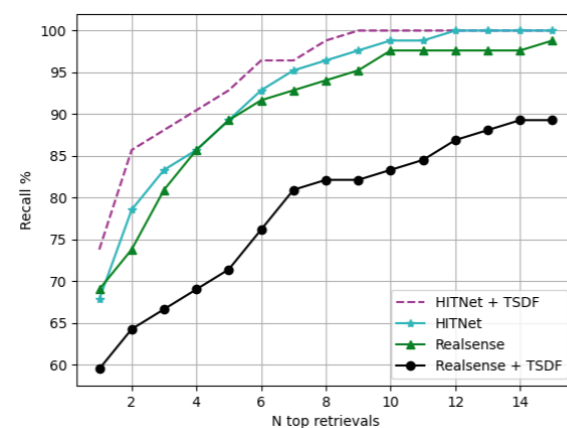


Figure 5: Recall results for querying 111 descriptors from 37 distinct shapes in arbitrary poses.

5 CONCLUSIONS AND OUTLOOK

By integrating the 3D process, the recognition accuracy can be improved while maintaining the performance of the system. According to a first estimation, a precision of > 95 % on thousands of objects can be achieved. With the now established option of parallelization and optimization of several methods, the recognition quality can be further improved in the future. The underlying data and fusion quality is a decisive factor here. The necessary data and fusion quality can be improved by targeting data pre-processing and consolidation. The spectrum of optimum data quality can vary depending on the application and part spectrum. This is largely dependent on the respective part spectrum to be acquired and the performance of the selected ML processes. The representative combination of HITNet fusion with TSDf 3D reconstruction and a point cloud descriptor like PointNetVLAD offers the greatest potential for further development at the current state of knowledge. Further topics and development paths for the field application which offer themselves in the perspective are among other things:

- The scalability of the system by the simple integration and comparison of the performance of further single and / or multi-step ML procedures on the basis of existing or additional data. (e.g., RGB based deep learning techniques)
- Investigate and expand the spectrum of the data space and data preprocessing from simple to complex for further use cases (targeted control of complexity).
- Cluster-specific narrowing of the search space (this feature provides significant improvements in performance and recognition quality)
- Transfer and extension of the procedures from the previous stationary concept to optional mobile components (tablet, smartphone or similar)
- Technical scalability through alternative sensor technologies and concepts, variable server-client architectures, graphical communication concepts, etc.
- Development of representative integration paths for specific industrial use cases.

REFERENCES

- [1] H. Hirschmüller, „Stereo-processing by semiglobal matching and mutual information.,“ 2008.
- [2] A. Kendall, H. Martirosyan, S. Dasgupta, P. Henry, R. Kennedy, A. Bachrach and A. Bry, „End-to-end learning of geometry and context for deep stereo regression.,“ in Proceedings of the IEEE International Conference on Computer Vision, 2017.
- [3] J. Pang, W. Sun, J. Ren, C. Yang and Q. Yan, „Cascade residual learning: A two-stage convolutional neural network for stereo matching.,“ In International Conference on Computer Vision-Workshop on Geometry Meets Deep Learning, 2017.
- [4] J.-R. Chang and Y.-S. Chen, „Pyramid stereo matching network.,“ in In IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [5] Z. Liang, Y. Feng, Y. Guo, H. Liu, L. Qiao, W. Chen, L. Zhou and J. Zhang, „Learning for disparity estimation through feature constancy.,“ 2017.
- [6] M. Bleyer, C. Rhemann and C. Rother, „Patchmatch stereo-stereo matching with slanted support windows.,“ in British Machine Vision Conference, 2011.
- [7] G. Yang, J. Manela, M. Happold and D. Ramanan, „Hierarchical deep stereo matching on high-resolution images.,“ in IEEE Conference on Computer Vision and Pattern Recognition, 2019.
- [8] F. Zhang, V. Prisacariu, R. Yang and P. H. Torr, „Ga-net: Guided aggregation net for end-to-end stereo matching.,“ in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
- [9] V. Tankovich, C. Häne, S. Fanello, Y. Zhang, S. Izadi and S. Bouaziz, „HITnet: Hierarchical iterative file refinement network for real-time stereo matching.,“ arXiv preprint arXiv:2007.12140, 2020.
- [10] Apple Inc., „Introducing Object Capture.,“ [Online]. Available: <https://developer.apple.com/augmented-reality/object-capture/>. [Accessed 31 10 2023].
- [11] M. Bookhahn, F. Essig, T. Lorenz, F. Voit and F. Neumann, „An Augmented Reality App supporting Service Technicians based on 3D Reconstruction and Object Recognition.,“ in GI VR / AR Workshop 2022, Stuttgart, 2022.
- [12] M. A. Uy and G. H. Lee, „Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition.,“ in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
- [13] C. R. Qi, H. Su, K. Mo and L. J. Guibas, „Pointnet: Deep learning on point sets for 3d classification and segmentation.,“ in Proc. Computer Vision and Pattern Recognition (CVPR), 2017.
- [14] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla and J. Sivic, „Netvlad: Cnn architecture for weakly supervised place recognition.,“ in In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [15] W. Zhang and C. Xiao, „PCAN: 3d attention map learning using contextual information for point cloud based retrieval.,“ in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.
- [16] C. R. Qi, L. Yi, H. Su and L. J. Guibas, „Pointnet++: Deep hierarchical feature learning on point sets in a metric space.,“ in Advances in neural information processing systems, 2017.
- [17] Z. Liu, S. Zhou, C. Suo, P. Yin, W. Chen, H. Wang, H. Li and Y.-H. Liu, „Lpd-net: 3d point cloud learning for large-scale place recognition and environment analysis.,“ in Proceedings of the IEEE International Conference on Computer Vision, 2019.
- [18] Y. Xia, Y. Xu, S. Li, R. Wang, J. Du, D. Cremers and U. Stilla, „SOE-Net: A Self-Attention and Orientation Encoding Network for Point Cloud based Place Recognition.,“ in Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [19] J. Komorowski, „Improving Point Cloud Based Place Recognition with Ranking-based Loss and Large Batch Training.,“ 26th International Conference on Pattern Recognition (ICPR), 2022.
- [20] Q.-Y. Zhou, J. Park and V. Koltun, „Open3D: A Modern Library for 3D Data Processing.,“ 2018. [Online]. Available: <http://www.open3d.org/wordpress/wp-content/paper.pdf>. [Accessed 30 10 2023].
- [21] CloudCompare, „CloudCompare 2.12 [GPL Software],“ 2022.

Segment Anything Model für Zero-Shot 3D Bounding Box Prediction

William Eppel¹, Marian Bookhahn¹, Frank Neumann¹

¹HTW – Hochschule für Technik und Wirtschaft Berlin
 Fachbereich 2: Ingenieurwissenschaften - Technik und Leben
 Studiengang Ingenieurinformatik
 Wilhelminenhofstr. 75a, 12459 Berlin
 eMail: frank.neumann@htw-berlin.de

Abstract. In der vorliegenden Arbeit beschreiben wir ein Verfahren, mit dem dreidimensionale Bounding Boxen für Bauteile in realen Scans ermittelt werden können. Dafür wurde die Vision-Transformer-basierte 2D-Segmentierungsmethode Segment Anything Model in Kombination mit einem Clusteringverfahren für die Rauschreduktion verwendet. Ein weiterer Teil unseres Verfahrens beschäftigt sich mit der automatisierten Orientierung der Bounding Boxen im Kontext ihrer Umgebung. Nachfolgend evaluieren wir die Effektivität unserer Technik in anwendungsorientierten Versuchen.

1 EINFÜHRUNG

Wenn Servicetechniker zur Wartung oder Reparatur von z.B. Heizungsanlagen, Produktionsanlagen oder Transportsystemen im Einsatz sind, treffen sie häufig auf Situationen, in denen die Artikelnummer eines dringend benötigten Ersatzteils nicht ad hoc ermittelt werden kann, weil die Anlagendokumentation nicht vorhanden oder nicht aktuell ist. In dieser Situation wäre es äußerst hilfreich, die Daten des benötigten Teils sofort zu bestimmen, z.B. indem man ein Bild oder eine 3D-Darstellung des Bauteils als Identifikationsmerkmal verwendet. Häufig sind die zu ersetzenden Teile jedoch entweder eingebaut und nicht vollständig sichtbar, stark verschlissen oder teilweise zerstört.

Das Forschungsprojekt SparePartAssist (gefördert durch das BMBF, Förderkennzeichen 16SV8511) adressiert dieses Szenario und zielt darauf ab, Servicetechniker zu unterstützen, indem das benötigte Ersatzteil auf Basis eines 3D-Scans der sichtbaren Teilegeometrie identifiziert wird. Anhand der von der SparePartAssist-App ermittelten Artikelnummer kann der Servicetechniker anschließend das passende Teil bestellen. Die Eingabedaten für den Objekterkennungsprozess werden von der Kamera, dem LiDAR-Sensor und der IMU des iPad Pro oder iPhone Pro geliefert. Wie in Abbildung 1 dargestellt, werden diese Eingabedaten anschließend durch den SLAM-Algorithmus von ARKit verarbeitet, der für jedes Kamerabild die zugehörige Kameraposition liefert. Diese Kamerapositionen und die Tiefenbilder bilden die Eingabedaten für eine TSDF-basierte 3D-Rekonstruktion, die eine Mesh-Darstellung für die akkumulierten Tiefenbilder generiert. Durch die Benutzeroberfläche der App

wird der Benutzer durch den gesamten Prozess geführt und kann den Fortschritt der Netzerstellung als Vorschau sehen, um die beste Scanqualität sicherzustellen. Nach der Fertigstellung des Scans kann der Benutzer den gewünschten Teil des Scans mit einer Bounding Box auswählen, um das gewählte Ersatzteil von der restlichen Szene freizustellen.

Auf der Grundlage des ausgewählten Netzes wird ein Merkmalsdeskriptor berechnet, der die Eingabedaten für eine geometrische Ähnlichkeitssuche bildet, die mit verdeckten Bauteilen umgehen kann.

Ein kritischer Faktor für SAMs Erfolg ist der umfangreiche Trainingsdatensatz SA-1B, bestehend aus 11 Millionen hochauflösenden Bildern und 1,1 Milliarden Segmentierungsmasken. Dieser beinhaltet deutlich mehr Daten als bisherige Datensätze, was SAM ermöglicht, in vielen Fällen bessere Segmentierungsergebnisse zu liefern als herkömmliche Modelle [3].

Des Weiteren ist im Rahmen dieser Arbeit das Clustering von hoher Bedeutung. Clustering ist ein Datenanalyseverfahren dessen Ziel es ist, Daten basierend auf der Ähnlichkeit ihrer Eigenschaften in sogenannte Cluster zu gruppieren [5] [6] [7]. Dabei sollen die Daten innerhalb

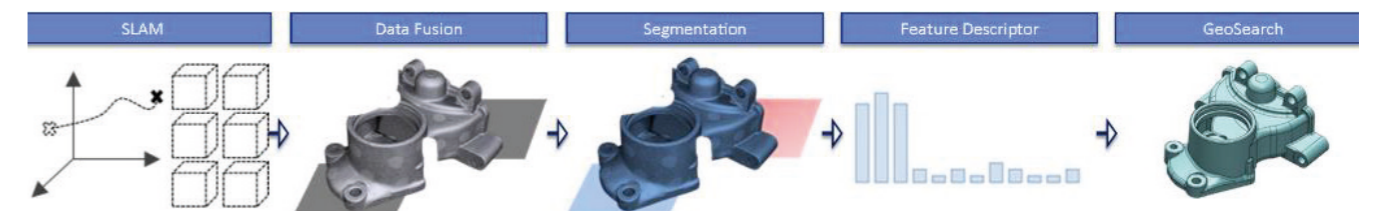


Abbildung 1: Rekonstruktions- und Suchpipeline der SparePartAssist-App [1]

Der o.g. Schritt der manuellen Auswahl eines Bauteils per Bounding Box durch den Benutzer soll durch die automatische Erzeugung von Vorschlägen basierend auf Segmentierungen ersetzt werden.

2 STAND DER TECHNIK

Es gibt viele verschiedene Ansätze zur Segmentierung. Grundsätzlich können Segmentierungsverfahren aber in zwei Klassen unterteilt werden, die klassischen Segmentierungsverfahren und gelernte Segmentierungsverfahren [1] [2].

Mit der fortschreitenden Entwicklung von Technologie haben Segmentierungsverfahren, die auf maschinellem Lernen basieren, in den vergangenen Jahren zunehmend an Bedeutung gewonnen. Diese weisen eine bemerkenswerte Leistungsverbesserung gegenüber den klassischen Verfahren auf [1] [2].

Ein State-of-the-Art-Verfahren, welches im Rahmen dieser Arbeit verwendet wird, ist das Segment Anything Model. Das Segment Anything Model (SAM) ist ein Foundation Modell für die Zero-Shot Bildsegmentierung [3]. Als Foundation Modell wird es mithilfe selbstüberwachter Lernmethoden und einer enormen Menge von Daten trainiert und kann Muster und Zusammenhänge in neuen, bisher nicht gesehenen Daten erkennen [4]. SAM hat die Fähigkeit, auf Basis eines Bildes und einer Segmentierungsaufforderungsgültige Segmentierungsmasken zurückzugeben [3].

eines Clusters eine hohe Intracluster-Homogenität und die Daten unterschiedlicher Cluster zueinander eine möglichst hohe Intercluster-Heterogenität aufweisen [8] [6] [7].

Es gibt eine Vielzahl verschiedener Clusteringalgorithmen, die unterschiedliche Ansätze verfolgen. Die Wahl des Verfahrens hängt dabei vom Anwendungsfall und den Daten ab. Im Rahmen dieser Arbeit fokussieren wir uns auf das Clustering von räumlichen Daten. In diesem Zusammenhang werden typischerweise die Verfahren DBSCAN [9], STING [10], Wavecluster [11] und CLARANS [12] verwendet [13].

3 PROBLEMANALYSE

Die automatische Generierung von 3D Bounding Box Vorschlägen stellt eine komplexe Aufgabe dar, bei der verschiedene Aspekte berücksichtigt werden müssen. Die Segmentierung spielt eine entscheidende Rolle, da sie das Zielobjekt vom restlichen Szeneninhalt trennt. Wenn die Segmentierung nicht sorgfältig durchgeführt wird, könnte die resultierende Bounding Box falsche oder irrelevante Informationen enthalten, was die Genauigkeit der Ersatzteilidentifikation beeinträchtigt.

Das Clustering ist ebenso wichtig, da es hilft, Störungen aus den segmentierten Daten zu entfernen. Ohne ein effektives Clustering könnten die Daten, wie in Abbildung 2 dargestellt, durch Rauschen verfälscht werden,

was zu weniger zuverlässigen Bounding Box Vorschlägen führt.

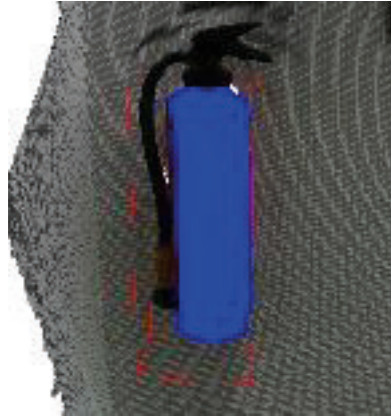


Abbildung 2: Einfluss von Rauschen auf die Qualität der Bounding-Box-Vorschläge

Schließlich ist die Anpassung und richtige Ausrichtung der Bounding Box entscheidend. Sie muss genau um das Zielobjekt passen und korrekt ausgerichtet sein, um unnötige Teile, wie z.B. den Untergrund, auszuschließen und nur das relevante Objekt einzuschließen. *Abbildung 3* veranschaulicht die Auswirkung einer nicht präzise ausgerichteten Bounding Box.

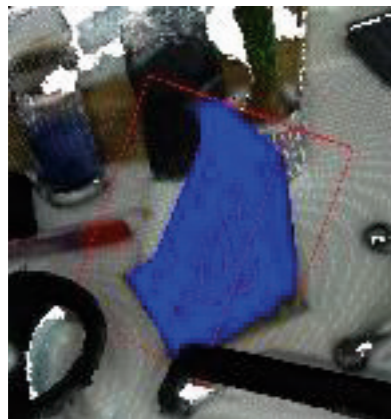


Abbildung 3: Unpräzise Bounding Box, die zu Inklusion irrelevanter Objektteile führt

Die Berücksichtigung dieser Faktoren bildet die Grundlage für die Definition der funktionalen und nicht-funktionalen Anforderungen. Diese Anforderungen sind entscheidend, um ein System zu entwickeln, welches in der Lage ist, zuverlässige und präzise 3D Bounding Box Vorschläge automatisch zu generieren.

Zu Beginn ist der Import und die Visualisierung eines Meshes wichtig, da dies die Basis für die nachfolgenden Schritte schafft. Weiterhin sollte es möglich sein, ein Objekt auszuwählen und aus dem Mesh herauszu-

segmentieren, da dieses den Bereich für die Bounding Box Prediction definiert. Ein zusätzlicher Schritt ist die Rauschreduktion durch Clustering, welche die Genauigkeit der Segmentierung und letztlich der Bounding Box Prediction verbessert. Die Generierung und Platzierung der Bounding Box bilden einen weiteren zentralen Prozessschritt. In Bezug auf die Optimierung der Bounding Box sollte die Bounding Box ein minimales Volumen aufweisen und parallel zur Ebene, auf der das ausgewählte Objekt platziert ist, ausgerichtet sein. Für Szenarien mit unvollständiger Segmentierung sollte die Möglichkeit bestehen, eine Verbesserung der Ergebnisse durch eine erneute Segmentierung zu erzielen. Eine zusätzliche Anforderung ist die farbliche Markierung der Vertices des sich in der Bounding Box befindlichen Objekts, was die visuelle Unterscheidung und Analyse erleichtert. Abschließend sollte die Möglichkeit bestehen, das segmentierte, rauschfreie Objekt zu exportieren, um die Weiterverarbeitung und Nutzung der Ergebnisse zu ermöglichen.

4 METHODENAUSWAHL

In diesem Kapitel erläutern wir detailliert den Auswahlprozess der Verfahren, die für die Entwicklung der vorgestellten Methode verwendet wurden. Verschiedene Auswahlkriterien wurden definiert und angewendet, um neben dem Segment Anything Model für die Segmentierung, welches sich in der durchgeführten Untersuchung als das Verfahren mit der besten Segmentierungsqualität erwies [14] [15], auch das passendste Verfahren für das Clustering auszuwählen. Die spezifischen Aspekte, nach denen die Verfahren beurteilt wurden, sind in Tabelle 1 aufgeführt.

Kriterium	Beschreibung
Effizienz	Das Verfahren soll große Datenmengen mit akzeptablem Zeit- und Speicheraufwand verarbeiten, um eine effiziente Nutzung der Anwendung zu gewährleisten.
Verfügbarkeit	Es soll bereits eine Implementierung des Verfahrens in einer Open Source Bibliothek vorhanden sein, da eine eigenständige Implementierung die Ressourcen, die für die Entwicklung der Anwendung vorhanden sind, übersteigt. Zudem stellt dies auch sicher, dass das Verfahren bereits getestet und optimiert ist.
Dokumentation	Es soll eine ausführliche und verständliche Dokumentation der Verfahren vorhanden sein, da dies die Verwendung, Anpassung und Integration der Verfahren erleichtert.
Universalität	Es soll nur eine geringe Abhängigkeit von benutzerabhängigen Parametern vorhanden sein, um die gesamte Anwendung einfacher und intuitiver zu gestalten. Zudem soll dadurch eine höhere Generalisierbarkeit der Anwendung ermöglicht werden.
Punktwolkenform-Generalisierung	Das Verfahren soll in der Lage sein, Cluster unterschiedlicher Formen in den zu verarbeitenden Daten zu erkennen. Dadurch kann eine hohe Generalisierung erreicht werden, was die Verarbeitung unterschiedlicher Geometrien und Strukturen ermöglicht.
Datenbereinigung	Das Verfahren soll in der Lage sein, potenzielle Ausreißer und Rauschen zu erkennen und zu entfernen. Dies gewährleistet, dass die Cluster nicht verfälscht sind und eine hohe Qualität aufweisen.
Automatische Clusteranzahlerkennung	Das Verfahren soll in der Lage sein, die Datenpunkte, ohne die Festlegung der Clusteranzahl durch den Benutzer, zu clustern. Dies ermöglicht es, die natürlichen Strukturen in den Daten zu erkennen und diese entsprechend zu clustern.

Tabelle 1: Kriterien für den Vergleich der Clusteringverfahren

In Bezug auf die Clusteringverfahren wurden, die in Kapitel 2 bereits erwähnten, typischen Verfahren für das Clustering räumlicher Daten miteinander verglichen. Die Tabelle 2 veranschaulicht den Vergleich der in Betracht gezogenen Clusteringverfahren.

	Effizienz	Verfügbarkeit	Dokumentation	Universalität
DBSCAN	o	++	++	++
STING	+	-	-	++
Wavecluster	+	--	-	++
CLARANS	-	+	+	+

	Punktwolkenform Generalisierung	Daten- bereinigung	Automatische Clusteranzahler- kennung	/
DBSCAN	++	++	++	
STING	++	--	++	/
Wavecluster	++	+	++	/
CLARANS	-	+	--	/

Tabelle 2: Vergleich der in Betracht kommenden Verfahren

Es konnte festgestellt werden, dass es für die Verfahren STING und Wavecluster keine passende öffentlich verfügbare Implementierung gibt und diese Verfahren zudem das Kriterium Dokumentation nicht in ausreichendem Maße erfüllen. CLARANS erweist sich als ungeeignet, da es vor allem nicht in der Lage ist Cluster, ohne Angabe der Clusteranzahl, zu erkennen [6], was essenziell im Rahmen dieser Arbeit ist. DBSCAN erfüllt insgesamt alle aufgestellten Kriterien in gutem Maße. Trotz eher moderater Effizienz [13], überzeugt DBSCAN primär durch gute Fähigkeiten im Umgang mit Rauschen [16], automatischer Clusteranzahlerkennung sowie sehr umfangreicher Dokumentation und wurde deshalb als das geeignetste Verfahren ausgewählt.

Das "Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise" (DBSCAN) Verfahren ist ein dichte-basiertes Clusteringverfahren [9]. Dichte-basierte Clusteringverfahren sind besonders gut geeignet, um Cluster beliebiger Form zu erkennen. Sie verfolgen den Ansatz, Datenpunkte basierend auf Regionen mit hoher Dichte in Cluster zu gruppieren [13]. Die Höhe der Dichte wird dabei durch einen bestimmten Schwellenwert festgelegt [6] [17].

Der DBSCAN-Algorithmus beginnt mit einem zufälligen, noch nicht klassifizierten Punkt aus dem Datensatz und untersucht die Umgebung dieses Punktes basierend auf vorgegebenen Parametern: dem Radius (Eps) und der minimalen Anzahl von Punkten (MinPts), die in dieser Umgebung vorhanden sein müssen [9].

Der Algorithmus klassifiziert Punkte in drei verschiedene

Kategorien: Kernpunkte, Grenzpunkte und Rauschen. Ein Kernpunkt ist ein Punkt, der mindestens MinPts in seiner Eps-Umgebung hat. Alle Punkte, die innerhalb des Nachbarschaftsradius eines Kernpunkts liegen, werden demselben Cluster zugeordnet. Wenn ein Punkt weniger Nachbarschaftspunkte als ein festgelegter Grenzwert in seiner Nachbarschaft hat, aber in der Eps-Umgebung eines Kernpunkts liegt, wird er als Grenzpunkt klassifiziert und ebenfalls dem Cluster zugeordnet. Punkte, die weder als Kern- noch als Grenzpunkte klassifiziert werden, werden als Rauschen betrachtet [9].

Der algorithmische Prozess ist iterativ und wiederholt die Untersuchung und Klassifizierung von Punkten, bis alle Punkte in dem Datensatz verarbeitet wurden. Dadurch entstehen Cluster, die auf Dichten basieren, und der Algorithmus kann komplexe und nicht-konvexe Clusterstrukturen identifizieren und Rauschen effektiv handhaben. Nach Abschluss des algorithmischen Prozesses werden Cluster, die aus Kern- und Grenzpunkten bestehen, und das Rauschen zurückgegeben [9].

BOUNDING BOX ORIENTIERUNG

Minimale Bounding Boxen für Punktwolken lassen sich sehr gut über die Hauptkomponentenanalyse (englisch PCA: Principal Component Analyses) approximieren. Der zusätzlich geringe Rechenaufwand (lineare zeitliche Abhängigkeit) und die einfache Implementierung machen die PCA zu einer sehr gängigen Methode für diese Aufgabe. Sind die drei Hauptachsen mittels PCA bestimmt, lässt sich die Ausdehnung der Bounding Box einfach über die Extremwerte der Projektion der Punkte auf diese Achsen ermitteln. Es ist zu beachten, dass eine heterogene Dichteverteilung der Punkte einen ungewünschten Einfluss auf die ermittelten Hauptkomponenten haben kann. Daher empfiehlt sich ein Punktwolken-Samplingverfahren, welches eine möglichst homogene Verteilung der gesampelten Punkte entlang der Oberfläche des Objektes gewährleistet.

Die so ermittelten Bounding Boxen sind häufig dennoch suboptimal für das Freistellen von Objekten in einer Szene. In vielen Szenarien sind die Objekte bedingt durch Wände, Böden oder Tische orientiert, sodass es sinnvoll erscheint, auch die Bounding Box daran zu orientieren. Im Folgenden soll eine Methode beschrieben werden, welche eine minimale Bounding Box unter der Vorgabe approximiert, dass eine Hauptachse durch den Normalenvektor \hat{n}_E der größten Fläche in der Nachbarschaft vorgegeben wird. Die Kovarianzmatrix einer Punktwolke X sei gegeben durch

$$\Sigma = cov(X) \quad (1)$$

, dann lassen sich durch das Lösen der Gleichung

$$\det(\Sigma - \lambda I) = 0 \quad (2)$$

die Eigenwerte λ_i der Kovarianzmatrix Σ ermitteln, wobei I der Einheitsmatrix entspricht. Mit den Eigenwerten aus Gleichung 2, lässt sich nun der i -te Eigenvektor \hat{e}_i mit

$$(\Sigma - \lambda_i I) * \hat{e}_i = 0 \quad (3)$$

ermitteln, welche jeweils einer Hauptkomponente der PCA entspricht. Die Hauptidee bei der Bestimmung der 2 fehlenden Achsen der neu orientierten Bounding Box (eine ist bereits durch den Normalenvektor \hat{n}_E festgelegt) besteht darin, diejenige Hauptachse Einfluss auf die neue Orientierung der Bounding Box nehmen zu lassen, welche sich am stärksten vom Normalenvektor \hat{n}_E unterscheidet. Wir ermitteln also nun den Eigenvektor \hat{e}_k , der zu dem Normalenvektor \hat{n}_E ein Minimum für den Betrag des Skalarproduktes annimmt mit

$$k = \underset{k \in \{1,2,3\}}{\operatorname{argmin}} \|\hat{e}_k * \hat{n}_E\| \quad (4)$$

Die zwei fehlenden Achsen \hat{n}_1 und \hat{n}_2 der neu orientierten Bounding Box ergeben sich dann durch

$$\hat{n}_1 = \frac{\hat{n}_E \times \hat{e}_k}{\|\hat{n}_E \times \hat{e}_k\|} \quad (5)$$

und

$$\hat{n}_2 = \hat{n}_E \times \hat{n}_1 \quad (6)$$

. Die Ausdehnung der neu orientierten Bounding Box ergibt sich wiederum durch die Extremwerte der Projektionen der Punkte auf die neuen Achsen \hat{n}_E , \hat{n}_1 und \hat{n}_2 .

5 EXPERIMENTE UND EVALUIERUNG

In diesem Abschnitt werden die Ergebnisse präsentiert und diskutiert, die während der Evaluierung der Leistung unserer Methode erzielt wurden. Die Tests wurden in einer spezifisch konfigurierten experimentellen Umgebung durchgeführt, bestehend aus einer Hardwarekonfiguration, die ein Dell Precision 7670 mit einer Intel Core i7-12850HX CPU, 32 GB DDR5 RAM und einer NVIDIA GeForce RTX 3080 Ti Grafikkarte umfasste. Softwareseitig wurde Ubuntu 22.04 verwendet, unterstützt von der NVIDIA Grafiktreiber-Version 525.125.06 und CUDA-Version 12.0.

Für die Evaluierung kam ein Testdatensatz zum Einsatz, der mit einer Intel Real-Sense D415 Kamera erfasst wurde, wobei die Abbildung 4 einen Teil dieses Datensatzes zeigt, um einen Einblick in die Art der gesammelten Daten zu ermöglichen.



Abbildung 4: Repräsentativer Auszug der Meshes aus dem Testdatensatz

Die Anwendung selbst ist in drei Hauptkomponenten unterteilt: Segmentierung, Clustering und Bounding-Box-Generierung. Diese Komponenten fokussieren sich jeweils auf unterschiedliche Aspekte des Prozesses. Im Rahmen der Evaluierung wurden sie hinsichtlich der Qualität der Ergebnisse sowie Laufzeit und Speicherbedarf untersucht.

Die Segmentierungskomponente kommt zum Einsatz, sobald ein Objekt im Mesh durch Anklicken eines Punktes ausgewählt wird. Ihre primäre Funktion besteht darin, ein 2D-Abbild der Szenenansicht zu erstellen (RGB- und Tiefenbild) und dieses, zusammen mit den Koordinaten des ausgewählten Punktes, mithilfe des Segment Anything Models zu verarbeiten. Basierend auf den Ergebnissen des Segment Anything Models wird dann durch Rückprojektion eine Punktwolke generiert, die das segmentierte Objekt repräsentiert. Ein repräsentatives Ergebnis der Segmentierung durch das Segment Anything Model sowie die erzeugte Punktwolke sind in den Abbildungen 5 und 6 dargestellt.



Abbildung 5: Ergebnis der Segmentierung: Links: Auswahl des Objekts in der Szene; Mitte: Segmentierungsergebnis (RGB); Rechts: Segmentierungsergebnis (Tiefenbild)



Abbildung 6: Punktwolke des segmentierten Objekts

Anhand einer visuellen Überprüfung ist ermittelt worden, dass die Segmentierungskomponente gute Ergebnisse liefert. Die Komponente ist in der Lage, die in der Szene ausgewählten Objekte präzise zu segmentieren und neue Punktwolken für diese Objekte zu erstellen. Zudem weist die Komponente bei einer durchschnittlichen

Laufzeit von 8 Sekunden eine akzeptable Verarbeitungsdauer und mit 4637,5 Megabyte einen akzeptablen Speicherbedarf auf.

Die Komponente Clustering wird verwendet, um die Qualität der Segmentierungsergebnisse zu verbessern, indem sie dazu beiträgt, die erzeugte Punktwolke zu verfeinern und etwaiges Rauschen zu eliminieren. Nach der Verarbeitung der Eingangsdaten erzeugt diese Komponente eine überarbeitete Punktwolke, die eine genauere Repräsentation des segmentierten Objekts darstellt. Durch diese Verarbeitung wird die Geometrie der Punktwolke optimiert, sodass sie näher an der tatsächlichen

Geometrie des Objekts liegt. Dies ist ein kritischer Schritt, um die Genauigkeit der nachfolgend generierten Bounding Box zu gewährleisten. Die Abbildung 7 veranschaulicht ein repräsentatives Ergebnis dieser Komponente.



Abbildung 7: Rauschreduzierte Punktwolke als Ergebnis des Clusterings

Es konnte ermittelt werden, dass die Komponente Clustering gute Ergebnisse liefert. Die Komponente ist in der Lage, die zu verarbeitenden Punktwolken vom Rauschen zu befreien und saubere Punktwolken, bei einer akzeptablen Laufzeit und einem akzeptablen Speicherbedarf, zu erzeugen. Die Laufzeit der Komponente liegt bei durchschnittlich 0,67 Sekunden pro 101.795 Punkte. Der Speicherbedarf beläuft sich auf durchschnittlich 13,45 Megabyte.

Die dritte Komponente ist für die Generierung einer Bounding Box zuständig. Diese Komponente erstellt eine Bounding Box, die das Objekt mit minimalem Volumen effizient umschließt. Zudem wird die Ausrichtung der Bounding Box sorgfältig angepasst, sodass sie unter anderem parallel zur größten Ebene des Meshes positioniert ist. Diese spezifische Ausrichtung sorgt dafür, dass unerwünschte und irrelevante Daten, wie der Untergrund, auf dem das Objekt platziert ist, von der Bounding Box ausgeschlossen werden, wodurch die Relevanz und Präzision der eingefangenen Daten verbessert wird. Die Abbildung 8 stellt repräsentativ die Ergebnisse dieser Komponente dar.

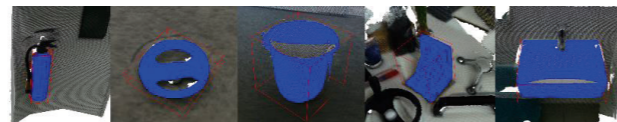


Abbildung 8: Ergebnisse der Bounding-Box-Generierung

Durch Tests konnte festgestellt werden, dass die Komponente gute Ergebnisse liefert. Die Komponente ist in der Lage, Bounding Boxen für die segmentierten Objekte zu erstellen, die überwiegend angemessen ausgerichtet sind und ein möglichst minimales Volumen aufweisen. Die Laufzeit der Komponente liegt bei durchschnittlich 0,24 Sekunden. Der Speicherbedarf beläuft sich auf durchschnittlich 38,5 Megabyte.

In der Abbildung 9 ist zudem dargestellt, wie bei nicht vollständiger Segmentierung eines Objekts die erzeugte Bounding Box durch erneute Auswahl eines anderen Punktes automatisch erweitert werden kann.



Abbildung 9: Erweiterung der Bounding Box: Links: Auswahl eines Punktes für den ersten Prozessdurchlauf; Mitte: Auswahl eines Punktes für den zweiten Prozessdurchlauf; Rechts: angepasste Bounding Box

Durch die Evaluierung der Gesamtperformanz der Methode wurde ermittelt, dass unsere Methode überwiegend gute Ergebnisse liefert. Die Anwendung ist in der Lage ein gewähltes Objekt einer 3D-Szene zu segmentieren sowie eine ausgerichtete minimale Bounding Box für das Objekt zu erstellen und in der Szene zu platzieren. Zudem ist anzumerken, dass die Anwendung auch mit nicht perfekten Scandaten umgehen kann, dabei gute Ergebnisse liefert und somit gut für den realen Einsatz geeignet ist.

Unter Berücksichtigung von zusätzlichem Downsampling vor dem Clustering beträgt die Verarbeitungsdauer eines Prozessdurchlaufs der Methode durchschnittlich 17,2 Sekunden, und der Speicherbedarf beläuft sich auf durchschnittlich 5454,9 Megabyte. Beide Parameter liegen im akzeptablen Bereich, vorausgesetzt, es wird eine angemessene Hardware-Konfiguration verwendet.

6 ZUSAMMENFASSUNG UND AUSBLICK

Die vorliegende Arbeit beschreibt eine Methode, mit der man initiale Bounding Boxen für fusionierte Tiefenbilder ermitteln kann. Sie bietet einen Überblick über Themenbereiche wie Segmentierung und Clustering anhand von bestimmten Verfahren, die bei der Anwendungsentwicklung verwendet wurden.

Insgesamt konnten die aufgestellten Anforderungen in gutem Maße erfüllt werden. Die Ergebnisse der Evaluierung haben gezeigt, dass die Anwendung gut geeignet ist, in dreidimensional dargestellten Szenen bestimmte Objekte zu erkennen, zu segmentieren und Bounding Boxen für diese zu erstellen. Aufgrund der Verwendung von maschinellem Lernen, überzeugen die Komponenten durch hohe Qualität der Ergebnisse.

Zudem ist die Benutzeroberfläche der Anwendung intuitiv und ermöglicht es auch Nutzern ohne spezielles Fachwissen diese erfolgreich zu bedienen.

Durch den hohen Generalisierungsgrad ist die Anwendung zudem fähig nicht nur im industriellen Kontext des Projektes SparePartAssist bestimmte Prozesse zu optimieren, sondern kann allgemein in allen Bereichen

angewendet werden, in denen eine effiziente und präzise Ermittlung von Bounding Boxen und die damit verbundene Segmentierung von dreidimensionalen Daten erwünscht ist.

Während die Anwendung bereits gute Resultate liefert, können zukünftige Forschung und neue Entwicklungen in den Bereichen der Bildverarbeitung und 3D-Computergrafik diese weiter optimieren. Neue Segmentierungsverfahren für dreidimensionale Daten, die vergleichbare oder bessere Ergebnisse erzielen als SAM, können die Effizienz und Effektivität der Anwendung weiter verbessern. Darüber hinaus könnte die Anwendung erweitert werden, indem die Möglichkeit der Erzeugung mehrerer Begrenzungsquader für unterschiedliche Objekte gleichzeitig geboten wird. Zusätzlich könnte die Anwendung durch Implementierung von GPU-Beschleunigung optimiert werden. Abschließend wäre es zudem interessant, die Anwendung in weiteren verschiedenen industriellen und nicht industriellen Kontexten zu testen, um ihre Vielseitigkeit eingehender zu evaluieren.

LITERATURVERZEICHNIS

- [1] S. Ghosh et al., „Understanding deep learning techniques for image segmentation.“ ACM computing surveys (CSUR), Bd. 52, Nr. 4, pp. 1-35, 2019.
- [2] M. Jena et al., „A survey on applications of machine learning techniques for medical image segmentation.“ International Journal of Engineering & Technology, Bd. 7, Nr. 4, pp. 4489-4495, 2018.
- [3] A. Kirillov et al., „Segment anything.“ 5 April 2023. [Online]. Available: <https://arxiv.org/abs/2304.02643>. [Zugriff am 24 August 2023].
- [4] R. Bommasani et al., „On the Opportunities and Risks of Foundation Models.“ 16 August 2021. [Online]. Available: <https://arxiv.org/abs/2108.07258>. [Zugriff am 25 August 2023].
- [5] E. A. Gupta et al., „Research paper on cluster techniques of data variations.“ International Journal of Advance Technology & Engineering Research, Bd. 1, Nr. 1, pp. 39-47, 2011.
- [6] G. J. Oyewole und G. A. Thopil, „Data clustering: Application and trends.“ Artificial Intelligence Review, Bd. 56, Nr. 7, pp. 6439-6475, 2023.
- [7] A. Novoselsky und E. Kagan, „An introduction to cluster analysis.“ 2021.
- [8] S. Bano und N. Khan, „A Survey of Data Clustering Methods.“ International Journal of Advanced Science and Technology, Bd. 113, 2018.
- [9] M. Ester et al., „A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.“ Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp. 226-231, 1996.
- [10] W. Wang et al., „STING: A statistical information grid approach to spatial data mining.“ VLDB, Bd. 97, pp. 186-195, 1997.
- [11] G. Sheikholeslami et al., „Wavecluster: A multi-resolution clustering approach for very large spatial databases.“ VLDB, Bd. 98, pp. 428-439, 1998.
- [12] R. T. Ng und J. Han, „CLARANS: A method for clustering objects for spatial data mining.“ IEEE transactions on knowledge and data engineering, Bd. 14, Nr. 5, pp. 1003-1016, 2002.
- [13] D. Xu und Y. Tian, „A Comprehensive Survey of Clustering Algorithms.“ Annals of Data Science, Bd. 2, 2015.
- [14] L. Xu et al., „MCTformer+: Multi-Class Token Transformer for Weakly Supervised Semantic Segmentation.“ 06 August 2023. [Online]. Available: <https://arxiv.org/abs/2308.03005>. [Zugriff am 04 September 2023].
- [15] W. Sun et al., „An Alternative to WSSS? An Empirical Study of the

Segment Anything Model (SAM) on Weakly-Supervised Semantic Segmentation Problems.," 02 Mai 2023. [Online]. Available: <https://arxiv.org/abs/2305.01586>. [Zugriff am 15 September 2023].

[16] M. Wegmann et al., „A review of systematic selection of clustering algorithms and their evaluation," 24 Juni 2021. [Online]. Available: <https://arxiv.org/abs/2106.12792>. [Zugriff am 27 August 2023].

[17] S. Roy und D. K. Bhattacharyya, „Data Clustering Techniques- A Review," 2007.

Informationsaustausch zwischen einer MixedReality-Anwendung für Stadtplanungsprozesse und BIM-Systemen mit dem Standard BCF

Angelika Beuster¹, Margitta Pries¹, Ute Wagner¹, Christoph Holtmann², Johann Habakuk Israel², Thomas Jung², Sebastian Keppler²

¹Berliner Hochschule für Technik
Fachbereich II - Mathematik, Physik, Chemie
Luxemburger Str. 10, D-13353 Berlin
<http://www.bht-berlin.de>

²HTW – Hochschule für Technik und Wirtschaft Berlin
Fachbereich 4: Informatik, Kommunikation und Wirtschaft
Wilhelminenhofstr. 75a, 12459 Berlin
[http:// www. htw-berlin.de/](http://www.htw-berlin.de/)

Abstract. Die Integration der verschiedenen Modellierungsarten von GIS und BIM ist für die Digitalisierung von Prozessen im Bauwesen von hoher Bedeutung. In diesem Beitrag wird ein Weg vorgestellt, Daten aus einem der GIS-Domäne zugeordneten System zur Bürger:innenbeteiligungen an demokratischen Entscheidungen in der Stadtentwicklung in den Bauplanungsprozess zurückzuführen. Hierzu wurde der internationale Open Standard BCF verwendet der zum Austausch von Informationen zwischen BIM-Systemen dient.

1 EINFÜHRUNG

Das BMBF-Projekt INSPIRER (Partizipation in Stadtplanungsprozessen in virtuellen und realen Räumen) [1] befasst sich mit der benutzer:innenzentrierten Entwicklung einer Anwendung mit AR¹- und VR²-Komponenten, die Bürger:innen zur Beteiligung an demokratischen Entscheidungsprozessen in der Stadtplanung anregen soll. Durch die Kombination von VR und AR sowie die Ergänzung um Multi-User- und weitere MR³-Technologien, kann die realitätsnahe Darstellung verbessert und zudem eine größere Unabhängigkeit bei den Bürger:innenbeteiligungsprozessen erreicht werden. Hierüber sollen die Teilhabe an der Stadtentwicklung erleichtert und inklusive Entscheidungsprozesse gefördert werden. Durch ein intuitiv nutzbares UI⁴ und asynchrone Interaktionen wird ein immersives Erleben virtueller Planungsstände ermöglicht, welche sich in das bestehende Stadtbild integrieren. Das System besteht aus zwei Client-Komponenten: eine AR-App erlaubt die Interaktion an verteilten Standorten im Planungsgebiet, während eine VR-App die standortunabhängige Beteiligung in einer virtuellen Stadt-Rekonstruktion ermöglicht. Über ein noch in Entwicklung befindliches innovatives 6-DOF⁵-Trackingverfahren sollen mobile Outdoor-AR-Anwendungen auf Basis von 3D-Punktwolken realisiert werden. Durch dieses Verfahren soll die

¹ AR: Augmented Reality

² VR: Virtual Reality

³ MR: Mixed Reality

⁴ UI: User Interface

⁵ DOF: Degree of Freedom

Genauigkeit gegenüber bestehenden Tracking-Lösungen verbessert werden, wodurch die Augmentierung virtueller Objekte im Stadtraum mit höherer Präzision möglich würde. Darüber hinaus werden die Benutzungsschnittstellen der AR- und VR-Anwendungen möglichst ähnlich umgesetzt, um eine einheitliche Bedienbarkeit zu gewährleisten und den Lernaufwand bei möglichen Wechselschritten zwischen den Plattformen gering zu halten. Der INSPIRER-Demonstrator verarbeitet GIS⁶-Daten. Im BIM⁷-Format vorliegende Planungsdaten werden daher vorab in ein gängiges GIS-Format übertragen. Das im INSPIRER-System erfasste Feedback der Bürger:innen soll im weiteren Stadtplanungsprozess berücksichtigt werden. Dazu ist es erforderlich, dass diese Informationen in einer geeigneten Form in BIM-Systemen integriert werden können. Trotz zahlreicher Berührungspunkte der beiden Domänen BIM und GIS besteht zwischen beiden Bereichen nach wie vor eine Trennung, die unter anderem durch Unterschiede in der Modellierung, der eingesetzten Software und der Verschiedenartigkeit der offenen Standards (z. B. CityGML für GIS und IFC für BIM) sowie auch durch eine unterschiedliche Entwicklungshistorie bedingt ist [2], [3]. Die hohe Aktualität des Themengebiets GIS-BIM-Integration und der große diesbezügliche Handlungsbedarf wurden unter anderem in einem Workshop mit 140 Bauwirtschaftsexpert:innen im Februar 2022 bestätigt [4]. Im Rahmen des Projekts INSPIRER wird daher erforscht, wie ein Informationsaustausch zwischen dem System INSPIRER aus der GIS-Domäne und BIM-Systemen bestmöglich umgesetzt werden kann.

2 VERWENDETE METHODEN

Im INSPIRER-Projekt liegen die Schwerpunkte wie erwähnt auf der Implementierung einer konsistenten, intuitiv nutzbaren und inklusiven Benutzungsschnittstelle, auf dem Austausch von Inhalten zwischen den AR- und VR-Apps sowie der Entwicklung eines neuartigen Trackingverfahrens für die AR-Anwendung. Die Integration von 3D-Modellen, digitalen Zwillingen, Punktwolken-Scans der städtischen Umgebungen sowie von Bauvorschriften und Bebauungsplänen soll den Anwender:innen eine umfassende Sicht auf die städtische Umgebung und geplante Projekte ermöglichen. Die Funktionen der Apps wurden dazu in einem partizipativen Entwicklungsprozess mit dem Ziel konzipiert, die Anforderungen von Bürger:innen unterschiedlichster Profile zu erfüllen. Die Beteiligung erfolgt durch Workshops, Interviews und Umfragen, um

sowohl quantitatives als auch qualitatives Feedback zu sammeln. Die Evaluation früher Entwicklungsstände der Demonstratoren und einzelner Funktionen ermöglichte das rasche Erkennen von Problemen und Erweiterungswünschen. Die sich während des iterativen Entwicklungsverlaufs aus dem Nutzer:innenfeedback regelmäßig verändernden Anforderungen als auch der technologischer Fortschritt im Projektverlauf erfordern einen agilen und modularen Entwicklungsansatz sowohl beim Design als auch bei der Implementierung der Anwendungen. Ein zentraler Bestandteil der VR- und AR-Anwendungen ist die Erfassung von Bürger:innenfeedback. Um diese Feedbackinformationen und weitere virtuelle Inhalte zwischen der AR- und der VR-Anwendung synchronisieren zu können, werden die Daten auf einem Planungsdatenserver abgelegt. Der INSPIRER-Planungsdatenserver hält neben den Anmerkungen und Änderungsvorschlägen die Daten der Planungsstände sowie virtuelle Stadtmodelldaten, die nach Bedarf zur AR-App und zur VR-App übertragen werden. Er bildet das zentrale Modul, um auf alle Entwurfsdaten einer Baustelle und deren geplanten Objekte zuzugreifen. Die von den Nutzer:innen erstellten Kommentare werden auf dem Server in Form von Annotationen, Skizzen und 3D-Objekten aus einem standardisierten Katalog abgespeichert. Das System stellt Änderungen dabei in Echtzeit im virtuellen Raum dar. Die Datenbank des INSPIRER-Planungsdatenservers wurde anhand eines gemeinsam im Rahmen des Projekts entwickelten Konzepts erstellt. Zu Kommentaren und Annotationen werden weitere notwendige Informationen, wie die Position des erstellten Feedbacks innerhalb der Szene (geografische Koordinaten), die Blickrichtung, der Erstellungszeitpunkt usw. erfasst. Diese Informationen sind besonders für die Übertragung in andere Systeme relevant. Damit die Feedbackinformationen der Bürger:innen den für die Bauplanung Verantwortlichen verfügbar gemacht werden können, müssen diese Daten mit den Daten des Bauprojekts, also insbesondere BIM-Modellen, auf direktem Weg verknüpft werden. Da sowohl Austauschformate aus dem CAD-Umfeld (z. B. STEP) als auch das IFC-Format die Speicherung von Annotationen ermöglichen, wurden zunächst darauf beruhende Ansätze analysiert. Allerdings zeigte sich sehr schnell, dass ein derartiger Lösungsansatz nicht zielführend ist. Sowohl bei der Konvertierung durch CAD- als auch durch BIM-Systeme werden derartige Informationen sehr häufig nicht vollständig übertragen.

Aus diesem Grund wurde der internationale Open Standard BCF⁸ genauer untersucht. Dieser dient dem

Austausch von Informationen und Nachrichten (Topics) zwischen BIM-Systemen und ist daher nicht im eigentlichen Sinne ein Datenaustauschformat. Man spricht auch vom „WhatsApp for BIM“. Aus der Trennung der Kommunikation vom Modell resultiert insbesondere eine Unabhängigkeit von der verwendeten Software. Unter anderem können über Kommentare zusätzliche Informationen hinterlegt werden, wobei durch die Angabe der entsprechenden IFC-GUID eine direkte Referenzierung mit einem IFC-Modell, z. B. mit dem in der IFC-Datei hinterlegten Projekt oder mit Bauelementen, möglich ist. Es gibt zwei grundlegende Möglichkeiten, BCF zu nutzen: über einen Webservice (bcfAPI) oder über einen dateibasierten Austausch (bcfXML). Im vorliegenden Anwendungsfall bietet sich der dateibasierte Austausch an. Eine BCF-Datei ist eine zip-Datei (.bcfzip) mit einer vorgegebenen Struktur [5], [6]. Pro Topic ist ein Unterordner mit einem bestimmten Ordnernamen anzulegen. Kernstück eines Topic-Ordners ist die XML-Datei markup.bcf, in der unter anderem die Kommentare bzw. Annotationen gespeichert werden können. Optional ist es außerdem möglich, Standortinformationen (XML-Datei viewpoint.bcfv) und einen „Schnappschuss“ (Bild-Datei snapshot.png) zu hinterlegen. Bis auf die png-Dateien für Schnappschüsse handelt es sich bei allen Dateien um XML-Dateien, die sehr einfach erstellt werden können. In praktisch allen gängigen BIM-Systemen stehen Funktionalitäten zum Einlesen von BCF-Dateien und zur Visualisierung der enthaltenen Informationen zur Verfügung.

3 ERGEBNISSE

Ein wichtiger Aspekt des Forschungsprojekts INSPIRER liegt wie oben erwähnt in der aktiven Bürger:innenbeteiligung an Stadtplanungsprozessen. In den VR- und AR-Anwendungen des Projektes ist es den Bürger:innen möglich, das geplante Bauvorhaben näher zu betrachten und Feedback in unterschiedlicher Form zu hinterlassen. Zum einen können Anmerkungen und Änderungswünsche in Form von Text-Annotationen eingetragen werden (Abb. 1).



Abb. 1. Annotation innerhalb der INSPIRER-VR-App

Feedback kann zudem über das Platzieren von vorgefertigten 3D-Objekten erfolgen, welche für das entsprechende Bauvorhaben genehmigt wurden (Abb. 2).



Abb. 2. Menü „Objekte platzieren“ der INSPIRER-VR-App

Zusätzlich besteht die Möglichkeit, freihändig 3D-Skizzen an gewünschter Stelle im Raum zu zeichnen, um beispielsweise kreative Ideen beizutragen (Abb. 3). Abb. 3.



Abb. 3. Menü „Skizzieren“ der INSPIRER-VR-App

⁶ GIS: Geoinformationssysteme

⁷ BIM: Building Information Modeling

⁸ BCF: BIM Collaboration Format

Die genannten Funktionen werden dabei sowohl in der AR- als auch in der VR-Anwendung angeboten. Dennoch gibt es Unterschiede zwischen den Plattformen: Der entwickelte AR-Smartphone-Demonstrator integriert die visuellen Inhalte und Konzepte der geplanten Stadtprojekte in das Kamerabild des Smartphones. Die Nutzenden können entsprechend Feedback anderer Bürger:innen und 3D-Visualisierungen der geplanten Projekte betrachten und erkunden, während sie sich durch die Stadt bewegen. Hierbei kommt der in Entwicklung befindliche Matching- und Tracking-Algorithmus auf Basis von Punktwolken zum Einsatz. Die Benutzungsoberfläche ist zudem auf Touch-Interaktionen sowie die Möglichkeiten von Smartphones angepasst, die bestimmten Einschränkungen beispielsweise hinsichtlich des Blickfeldes und der Nutzung der Hände unterliegen. Es werden aber auch in der Umgebung platzierte, dreidimensionale Oberflächen getestet und teils mit dem Konzept des passiven haptischen Feedbacks umgesetzt. Hierfür werden UI-Elemente beispielsweise auf Oberflächen realer Objekte, wie Stromkästen, platziert. Der VR-Client ermöglicht es Nutzenden, sich standortunabhängig durch eine virtuelle Repräsentation einer Stadt bzw. eines Ortes mit Bauprojekten zu bewegen. In diesem Fall wird die bestehende Umgebung in Form von virtuellen dreidimensionalen Gebäude- und Landschaftsmodellen dargestellt, in welche weitere Inhalte integriert werden. Die Interaktionen sind für die Steuerung mit VR-Controllern optimiert. Auch die Menüs und Interface-Elemente sind an die sichtfeldfüllende Darstellung angepasst. Die Interface-Strukturen bleiben weitestgehend zu denen des AR-Demonstrators konsistent. Von Nutzenden eingefügtes Feedback wird ortsgenau an den betreffenden geografischen Koordinaten platziert, sodass es nach der Synchronisation mit dem Planungsdatenserver in der AR-App exakt in der realen Umgebung visualisiert werden kann.

Im Rahmen des Forschungsprojektes wird exemplarisch demonstriert, wie in den VR- und AR-Anwendungen erfasstes Feedback unter Nutzung des BCF-Standards in verschiedene BIM-Systeme übertragen werden kann. (Abb. 4).

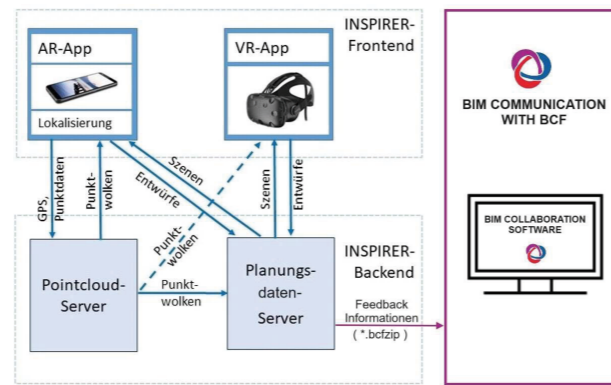


Abb. 4. Erweiterte INSPIRER-Systemarchitektur und Übertragung in BIM Software

Dazu werden auf dem Planungsdatenserver gespeicherte Feedbackinformationen automatisiert in ein BCF-Dokument übertragen. Die Annotationen werden in einer XML-Datei markup.bcf, deren Standortinformationen in der Datei viewpoint.bcf erfasst. Zusätzliches über normale Annotationen hinausgehendes Feedback, wie platzierte Objekte oder angefertigte Skizzen, kann per Screenshot gespeichert und ebenfalls übermittelt werden. Die Dateien werden zu einem einzigen, als BCF-Datei bezeichneten Dokument zusammengefasst, das in einem BIM-System geöffnet werden kann.

```

1 <?xml version="1.0" encoding="utf-8"?>
2 <Markup>
3 <!--<Topic-Guid="562991a0-7e24-4c04-9f9d-e8ec793b62c8"-TopicType="Issue"-TopicStatus="Active">
4 <!--<Title>Fenster ändern</Title>
5 <!--<Priority>Major</Priority>
6 <!--<Index>2</Index>
7 <!--<CreationDate>2023-05-21T06:32:16+00:00</CreationDate>
8 <!--<CreationAuthor>angelika.beuster@bht-berlin.de</CreationAuthor>
9 <!--<ModifiedDate>2023-05-21T06:32:16+00:00</ModifiedDate>
10 <!--<ModifiedAuthor>angelika.beuster@bht-berlin.de</ModifiedAuthor>
11 <!--<DueDate>2023-05-21T08:00:00+00:00</DueDate>
12 <!--<AssignedTo>angelika.beuster@bht-berlin.de</AssignedTo>
13 <!--<Description>Fenster sollte ein Sprossenfenster sein</Description>
14 <!--</Topic>
15 <!--<Comment-Guid="60c7f423-ae56-44ac-a375-75f6314e6a11">
16 <!--<Date>2023-05-21T06:32:16+00:00</Date>
17 <!--<Author>angelika.beuster@bht-berlin.de</Author>
18 <!--<Comment></Comment>
19 <!--<Viewpoint-Guid="d2a3e24d-9495-4c84-ad21-ca2e31428120">
20 <!--<ModifiedDate>2023-05-21T06:32:16+00:00</ModifiedDate>
21 <!--<ModifiedAuthor>angelika.beuster@bht-berlin.de</ModifiedAuthor>
22 <!--</Comment>
23 <!--<Viewpoints-Guid="d2a3e24d-9495-4c84-ad21-ca2e31428120">
24 <!--<Viewpoint>viewpoint.bcf</Viewpoint>
25 <!--<Snapshot>snapshot.png</Snapshot>
26 <!--</Viewpoints>
27 <!--</Markup>

```

Abb. 5. Codebeispiel einer markup.bcf Datei

Die XML-Datei markup.bcf (Abb. 5) enthält neben der ID und dem Typ des Topics, dem Bearbeitungsstatus, dem Titel und der Priorität weitere Angaben wie Ersteller:in, Erstellungs-, Änderungs- und Fälligkeitsdatum sowie eine genauere Beschreibung des Topics. Zusätzlich werden Kommentare, ebenfalls per eindeutiger ID, und deren Verfasser:in und Datum abgespeichert. Gleiches gilt für die Standortinformation. Genauere Angaben des Standorts wie Informationen zur Position und Richtung der Kamera als auch Schnittebenen lassen sich der XML-Datei viewpoint.bcfv entnehmen.

Für alle gängigen BIM-Systeme stehen Funktionalitäten bzw. Plugins für die Anzeige von BCF-Dateien zur Verfügung. Für genauere Testungen der Datenübertragung wurde der BIMcollab BCF MANAGER genutzt, der für Navisworks, Revit, Archicad, Tekla Structures und verschiedene AutoCAD Module verfügbar ist (Abb. 6). Mit Hilfe dieses Plugins werden die verschiedenen Topics, hier Issues genannt, in einer Übersicht dargestellt. Nach Auswahl eines dieser Issues wird die Position des aktuellen Modells so ausgerichtet, dass die relevante Stelle im Modell ersichtlich ist. Das Ansichtsfenster enthält neben Verfasser:in ebenso eine genauere Beschreibung und den aktuellen Bearbeitungsstatus des Issues. Zudem ist es hier möglich, das Issue zur Bearbeitung einer Person zuzuweisen und weitere Kommentare zu verfassen. Auch kann ein zusätzliches Issue erzeugt werden. Hierfür muss das Modell zunächst innerhalb der Software so gedreht werden, dass das neu zu erstellende Issue gut sichtbar ist. Dies ist erforderlich, da beim Erstellen des Issues automatisch ein Snapshot generiert wird. Dieser Snapshot kann dann zum optischen Hervorheben des Problems in einer simplen Zeichenanwendung innerhalb des BCF MANAGERS geöffnet und nachbearbeitet werden.

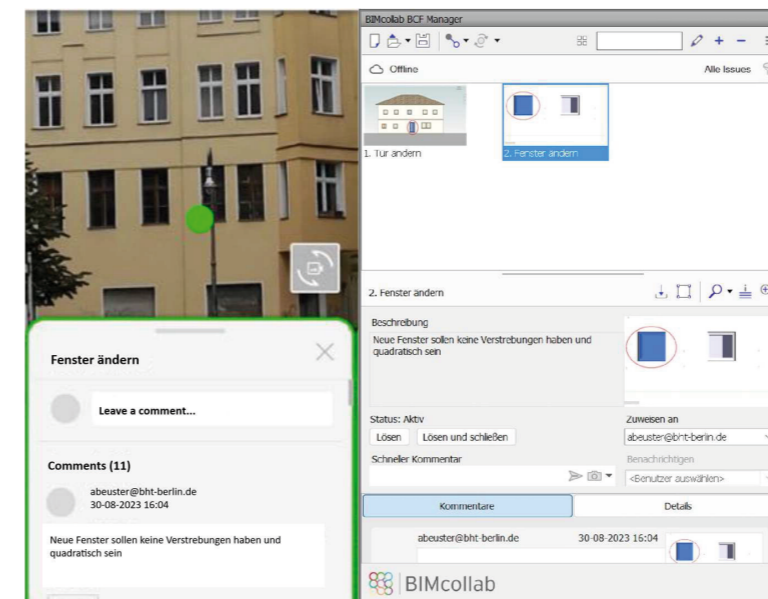


Abb. 6. Nutzer:innen Kommentar in INSPIRER AR-App (links), BCF-Kommentar in Autodesk Navisworks Manage 2023 im BIMcollab BCF MANAGER (rechts)

4 DISKUSSION

Bei der kontinuierlichen, verzahnten Entwicklung und Evaluation im Projekt INSPIRER hat sich gezeigt, dass eine erfolgreiche Einbeziehung von diversen Nutzer:innengruppen in Stadtplanungsprozesse mittels VR- und

AR-Systemen nur gelingen kann, wenn der Datenaustausch zwischen allen Beteiligten – den Architekt:innen und Landschaftsgestaltenden in den Planungsbüros, den Verantwortlichen in der kommunalen Verwaltung, den Entscheidenden in den entsprechenden Gremien und den Bürger:innen – schnell und unkompliziert erfolgt. Mit BIM steht dafür eine moderne, offene Basis mit einer großen Vielfalt an unterstützenden Software-Tools zur Verfügung. Darauf aufbauend bietet die hier aufgezeigte Vorgehensweise eine Option, Daten aus einem der GIS-Domäne zugeordneten System in den Planungsprozess zurückzuführen, und liefert damit einen Beitrag zur aktuell im Fokus stehenden GIS-BIM-Integration. Dabei wird jedoch vorausgesetzt, dass sämtliche Planungsdaten digital als 3D-Modell, vorzugsweise im IFC-Format verfügbar sind, was heutzutage in der Praxis noch selten der Fall ist. Die große Vielfalt, in welcher Art Informationen zum Planungsstand an den verschiedenen Stationen des Entwurfs-, Beteiligungs- und Entscheidungsprozesses vorliegen, bildet eine Herausforderung für die Entwicklung von Softwaretools, die die gesamte Planungsprozesskette unterstützen und mit intuitiv nutzbaren, ansprechenden UI-Komponenten eine digitalisierte, partizipative Stadtplanung gewährleisten können.

5 FAZIT

In diesem Beitrag wurde die Eignung von BCF für einen Informationsaustausch zwischen unterschiedlichen Domänen zugeordneten Systemen diskutiert und Umsetzungsmöglichkeiten anhand der Entwicklungen im INSPIRER-Projekt demonstriert. Die Formate und Verfahren bieten sich für das angestrebte Ziel an, einen Informationsaustausch über AR- und VR-Geräte hinweg zu gewährleisten und damit partizipative und kollaborative Stadtplanung zu ermöglichen. Die Übertragbarkeit der gesammelten Daten, Annotationen und Ideen in BIM-Systeme und weiterführende Planungsprozesse ist dabei entscheidend für die Umsetzung. Das BCF-Format stellt sich hierfür als besonders sinnvoll heraus, da es bei Bedarf über Plugins in praktisch alle gängigen BIM-Systeme eingelesen werden kann. Eine aktuelle Herausforderung im INSPIRER-Projekt besteht unter anderem darin, die nötigen Informationen so zu organisieren, dass diese bei der weiteren Nutzung sinnvoll angezeigt und weiterverwendet werden können.

LITERATURVERZEICHNIS

- [1] INSPIRER, „Projekt INSPIRER,“ [Online]. Available: <https://projekt-inspirer.de/>. [Zugriff am 11 09 2023].
- [2] H. Lehmkuhler, J. Kauer und R. Steinmann, BIM & GIS, Berlin: Wichmann, 2022.
- [3] TU Delft, „GeoBIM: Bridging the gap between Geo and BIM,“ [Online]. Available: <https://3d.bk.tudelft.nl/projects/geobim/>. [Zugriff am 18 07 2022].
- [4] buildingSMART, „BIM und GIS - erfolgreicher Kick-Off von buildingSMART,“ [Online]. Available: <https://www.buildingsmart.de/buildingsmart/aktuelles/bim-und-gis-erfolgreicher-kick-von-buildingsmart>. [Zugriff am 2022].
- [5] buildingSMART, „BCF-XML/Documentation at releade_3_0 buildingSMART/BCF-XML,“ [Online]. Available: https://github.com/BuildingSMART/BCF-XML/tree/release_3_0/Documentation. [Zugriff am 11 09 2023].
- [6] buildingSMART, „BIM Collaboration Format (BCF),“ [Online]. Available: <https://www.buildingsmart.org/standards/bsi-standards/bim-collaboration-format-bcf/>. [Zugriff am 26 10 2023].

Comparison of Digital Model with Model in Production for Quality Assurance Purposes

Bhumil Dhanani (Think3DDD), Amar Kumar Myadam (Think3DDD, 3D-Medico), Tino Jacobi (Think3DDD, 3D-Medico)

Think3DDD GbR,
Teutonenstraße 74,
12524 Berlin
www.think3ddd.de

Abstract. The use of 3D printing technology has rapidly expanded in the medical device industry, offering new opportunities for highly customized and complex structures. However, ensuring the quality of the final product is essential and has become a critical concern. To ensure reliability and accuracy, quality assurance methods are needed. This study's goal is to evaluate the suitability of digital models for use in 3D printing medical devices for quality control. The study involves a detailed analysis of the accuracy of digital models compared to the models produced in the manufacturing process, using Digital photographic measurements and analysis techniques. By focusing on the dimensional accuracy and surface finish of the digital and physical models, any differences and similarities will be identified to understand the extent of reliability and accuracy of the digital models. The research's conclusions have significant implications for the use of digital models in quality assurance in the 3D printing of medical devices. By identifying the strengths and limitations of digital models compared to physical models, this research will contribute to the development of more effective quality assurance practices, which can lead to higher quality and more reliable medical devices, improving patient outcomes. This study can assist in guiding future research in this area by offering insightful information about the usage of digital models for quality assurance.

Key Points: 3D Printing, Quality Assurance, FDM Printing, Comparison, Digital Photographic Measurements

1 INTRODUCTION

Early in the 1980s, the desire for quicker and more affordable solutions sparked the creation of 3D printing. For quick and affordable methods for creating prototypes for mass production. Charles Hull, a US citizen, was the inventor of the first 3D printing technique. He applied for his first patent for the manufacturing method stereolithography in 1984 and was successful in getting it in 1986 [1].

In recent years, 3D printing has received much interest due to its usefulness in producing prototypes before mass production. The rapid development of this technology has allowed 3D printers to compete with traditional factories in terms of both speed and quantity of form production. Accurate production is possible with additive manufacturing thanks to its use of digital scanning and dimension control, making it a desirable alternative for many companies. Because of the growing importance of 3D printing in healthcare, it is crucial to incorporate quality assurance (QA) processes from the start. In this study, a variety of techniques are used to assess how accurate a „3D-printed anatomical model“ is. Numerous methods can be used to take measurements., which generally include both physical and digital scans [2]. This study's main objective is to present a thorough strategy for creating a program for „quality assurance“ that might be applied to medical 3D printing applications. This research suggests using digital photographic measurement as a quick and cheap method for checking the accuracy and precision of 3D printing.

1.1 RESEARCH OBJECTIVES

This study's main objective is to assess the quality of 3D printed products by contrasting a digital model with a physical one. Python and slicing software will be used in the study to examine and contrast the data from the two models. The results of this investigation will advance the present body of knowledge and provide light on the possible advantages and disadvantages of employing digital models for quality assurance in 3D printing.

Objectives:

- To develop a quality assurance (QA) program for use in medical 3D printing applications using Python programming language and Slicing software.
- To explore various techniques for the quality assessment of a 3D-printed anatomical model's accuracy.
- To compare images between real images and generated images in slice using Python programming language.
- To validate the printed models by qualitative inspection and quantitative measurement.

2 SURVEY OF EXISTING STUDIES

This study will explore the current research and literature surrounding the comparison of digital models with models in production for quality assurance purposes in 3D printing, analyze the current state of research and technology in the 3D printing field, and explore how the comparison between digital and production models can be used to ensure quality assurance.

2.1 CONSIDERATIONS FOR FUSED FILAMENT 3D-PRINTING

Several initiatives have been launched to enhance 3D printers' dependability by using high-resolution photos [3]. Most of these initiatives have, however, used 3D printing techniques that need high-priced laser equipment for their precision. An explanation of the various built-in error detection features of the EOS INT M 270 Laser Beam Melting System, such as the monochrome CCD camera system, the tilt-and-shift lens that helps reduce perspective distortion, and the adjustable tube that allows to change the focal length of the system or the separation between the object and the lens. Furthermore, they used their pictures to examine the results of a powder bed AM process for defects, missing powder or low energy input, surface quality, and assessments

of component geometry on a microscopic scale. To provide high-resolution image-based measurements, it is necessary to calibrate the perspective utilizing their template matching method.

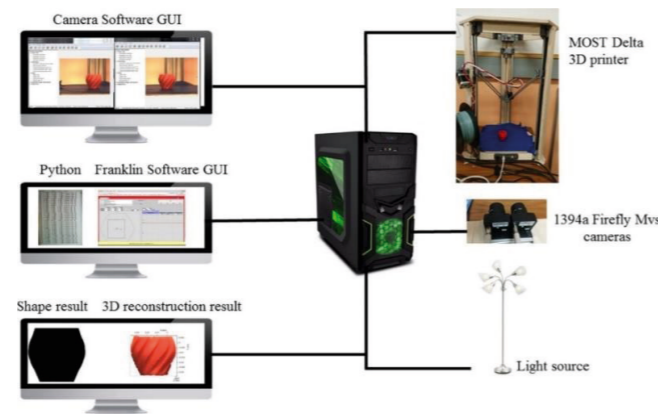


Figure 1: Setting up an optically monitored Delta printer [3]

High-resolution imagery and an improved setup with a proximity sensor are two methods for achieving more consistent results from 3D printing. Several proprietary computer vision-based commercial solutions are also available for luxurious printers. Few studies have been conducted on how to identify errors in desktop FFF-based 3D printing for prosumers. [4]. Commercial FFF printers often include a sensor that can detect whether the nozzles are blocked or when the filament has run out. 3D, a distributor of the Gigabot brand of open-source 3D printers, provides a filament detector as a premium add-on in response to customer complaints about print delays. Each of these procedures is designed to do away with a different possible failure mode [5].

It is possible to detect an error from a single camera setup by comparing the simulated image to the 2D form model after generating a 2D shape picture (Camera Image) of the 3D item. To generate an STL Image, one renders a 3D model in OpenSCAD and saves the result as an STL file (STL Image). Then, using STL tools, one plots all the data in the STL file along the x, y, and z axes to show the form of the rendered 3D model from various perspectives. When charting a model, the camera's perspective must be adjusted precisely so that it corresponds to the location of the object being seen. Thus, the correct viewing angle results in a PNG image of the STL images being stored on the XZ plane.

2.2 UNIT-PRODUCTION 3D PRINTING QUALITY ANALYSES USING FDM AND MSLA TECHNOLOGIES

Two distinct 3D printing processes were used. Fusion Deposition Modeling came first. Melt deposition is the foundation of FDM technology. The process involves applying the model and the support material via the nozzle in the head of the printer in alternating layers, as shown in the subsequent sectional views of the intended model.

Acrylonitrile-butadiene-styrene (ABS), polylactic acid (PLA), and polyethylene terephthalate (PET-G) are the most popular polymers utilized in FDM technology's fundamental fabrication processes [6]. Masked Stereolithography (MSLA) is used in this method to chemically cross-link monomers and oligomers. This procedure leads to the synthesis of polymers into a solid three-dimensional shape. Materials termed „resins“ are employed extensively in this technique. These polymers harden when exposed to heat. Additive manufacturing techniques like 3D printing are often used for single-item or low-volume production runs [7].

Individual requirements are met with the help of 3D printing technology, which has many useful applications. In recent years, 3D printers have grown more accessible and inexpensive enough for home usage. Any desired item may be printed by a user using 3D printing technology. If the user is missing a piece, they may simply print it out. The rising popularity of 3D printing prompted studies to compare the two main types of technology to see which would be better suited for mass production, FDM or MSLA. The printing time for an item using masked SLA technology is proportional to its height (the number of layers). This is the primary constraint on the validity of the study's findings and multiple sources attest to the fact that printing materials created using MSLA technology are more difficult to create [8]. This approach is more challenging and time-consuming for the ordinary user in terms of unit output since post-processes like cleaning the prints as well as their ultimate hardening with UV light are required.

3 METHODOLOGY

The methods used to collect data for the subject are described in this portion of the study. This section describes in depth the philosophy, methodology, design, data-collecting method, tools, and procedures used to accomplish the study's stated goals. The research

procedure and the reliability of the results are explained and evaluated in the methodology section. Ethical concerns that were made during the study are also discussed in this chapter. In this investigation, we employed a primary strategy that included quantitative techniques. The study set out to ensure the quality of 3D printed goods by contrasting the accuracy and dependability of a digital model with the model in production. How the study goal was accomplished and how the data was examined will be explained in great depth in the chapter devoted to methodology. Here, we also explore the study's weaknesses and potential future research approaches.

The methodology section of this work serves the purpose of providing a thorough description of the research approach and techniques employed to achieve the study's stated goals and objectives. Overarchingly, this research aims to evaluate the complete 3D printing process for its precision and accuracy to build a “quality assurance program for medical grade 3D printing equipment” [9]. The correctness of a 3D-printed anatomical model will be evaluated using a variety of methods, which will include measurements from different models.

3.1 TOOLS AND TECHNIQUES

Several tools and instruments were used in the study to gather the necessary data, which are mainly divided into two categories, software and scanning tools. To create a quality assurance comparison model or software, this research has heavily relied on the Python programming language technique and for the data gathering this study has relied on the Slicing software (CURA Ultimaker) [10]. These two tools/software have been the most two important pieces of technology that this study has utilized. Additionally, for the scanning of the 3D models, techniques like physical and digital photographic measurements, and photogrammetry were utilized, which further helped this study to determine the accuracy of the 3D-printed models. The computer was able to evaluate the accuracy and correctness of each stage of the 3D printing process by comparing the resulting three-dimensional model to that same initial QA phantom using image-based criteria. In addition, 3D objects were uploaded (In . STL file format) in Slicing Software and compared to real-world counterparts using Python programming language. The overall accuracy and precision of the 3D printing process were evaluated, and a QA program for 3D printing

of medical-grade equipment was developed, thanks in large part to the use of these tools and procedures.

4 CONSTRUCTIVE SPECIFICATIONS

The reader will receive an overview of several fundamental additive manufacturing specifications or parameters in this section, including both digital and physical models.

4.1.1 DESIGN PROCESS

To build the physical model, the design process used a computer model made in Blender that was then transformed into a format for 3D printing. Typically, this procedure entails several processes, such as choosing the best 3D printing technique, the right materials, and the right printing settings. The choice of the 3D printing method that best meets the needs of the created object is the first stage in the design process. Depending on the particular application, this can be Fused Deposition Modeling (FDM), Stereolithography (SLA), Selective Laser Sintering (SLS), or another technology [11]. FDM has been selected as the technique to be used to generate the 3D print model. The Company has already used that technique.

4.1.2 SLICING PARAMETERS

Slicing parameters refer to the configurations made in the slicing software (like CURA Ultimaker) to produce the instructions that the 3D printer will follow. Among other things, these settings include layer height, infill density, print speed, and temperature. There are numerous parameters associated with this program; however, we will focus on a few of the most crucial ones, including layer height, wall thickness, printing temperature, infill density, build plate temperature, printing speed, bottom layers, etc. Some of the most important slicing parameters are listed here [12].

Parameter	Value
Layer Height	0.2 mm
Wall Thickness	0.8 mm
Wall line count	2
Top and bottom thickness	0.8
Top layers	4
Bottom Layers	4
Infill density	20 %

Infill pattern	Cubic
Printing temperature	200 °C
Build plate temperature	55 °C
Print speed	60 mm/s
Retraction	Enabled
Fan Speed	100 %
Supports	No supports
Build plate adhesion type	Brim

Table 1: Key printing parameters for FDM printing technology (parameters available in CURA 5.1.1 Software for cube print)

4.2.1 3D PRINTER TYPES

A comparison of the 3D printing method and quality analyses of FDM and MSLA technologies was covered in the preceding literature review chapter. An FDM 3D printer was used to create the physical model. There are numerous varieties of 3D printers, each with special qualities and benefits.

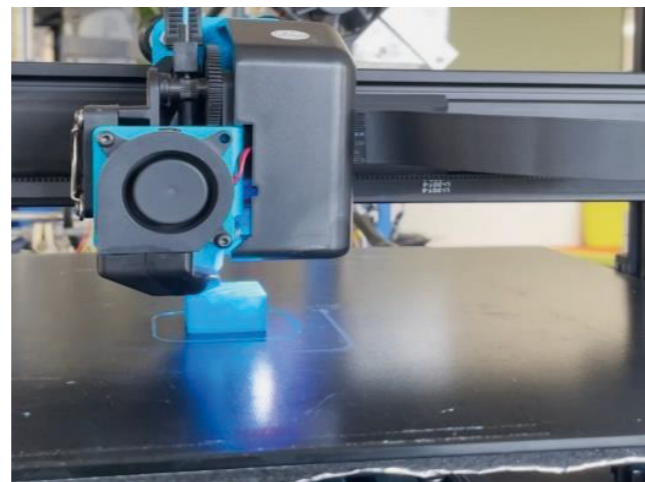


Figure 2: FDM 3D Printer

4.2.2 PRINTING MATERIAL

The printing material is used to create printed or actual things. There are various commonly used printing materials available on the market, but each material has a unique set of properties. Depending on the kind of printer and the intended application of the printed object, several printing materials can be used in 3D printing. ABS and PLA are common thermoplastics used in FDM printing, whereas resins are frequently used in SLA printing. The 3D printing process can also be utilized with other

materials, including metals, ceramics, and even food items. The printing substance was made of PLA filament. Using renewable materials like corn starch or sugarcane, PLA (polylactic acid) is a biodegradable thermoplastic. Due to its simplicity of use, lack of warping, and variety of colors and finishes, it is a common material for 3D printing [13].

4.2.3 PRINTING PROCESS

The printing process is divided into three parts pre-printing, printing, and post-printing. Pre-printing was done by design and slicing software. Then after printing was done by the FDM Printer, the last part of post-printing removed extra things from the object, and the surface finishing of the final product. All of them are a part of the printing process, however, the following steps will help you grasp it more clearly [14].

4.2.4 MODEL ACCURACY

A digital caliper was used to take several measurements on both the real model and the digital model to assess the accuracy. The printed model's width, height, and length as well as the holes' diameter and depth were all measured. Hence, to measure the accuracy of the model, the maximum error and measurement tolerance can be determined from ± 0.5 mm, and the maximum error for the orthosis is 0.15mm. This means that any deviation within ± 0.5 mm from the actual dimension is within the acceptable tolerance range. The standard deviation calculation is shown below:

For the orthosis's Deviation:

$$\begin{aligned} \text{Length} &= 78.59 - 78.50 = 0.09 \text{ mm} \\ \text{Width} &= 100.62 - 100.70 = -0.08 \text{ mm} \\ \text{Height} &= 245.75 - 245.90 = -0.15 \text{ mm} \end{aligned}$$

The Squared deviation for each dimension:

$$\begin{aligned} \text{Squared Deviation in length} &= (78.59 \text{ mm} - 78.50 \text{ mm})^2 = 0.0081 \text{ mm} \\ \text{Squared Deviation in width} &= (100.62 \text{ mm} - 100.70 \text{ mm})^2 = 0.0064 \text{ mm} \\ \text{Squared Deviation in height} &= (245.75 \text{ mm} - 245.90 \text{ mm})^2 = 0.0225 \text{ mm} \end{aligned}$$

The average of the squared deviations:

$$\begin{aligned} \text{Average Squared Deviation} &= (0.0081 \text{ mm} + 0.0064 \text{ mm} + 0.0225 \text{ mm}) / 3 \\ &= 0.0037 \text{ mm} / 3 \\ &= 0.0123 \text{ mm} \end{aligned}$$

The standard deviation:

$$\begin{aligned} \text{Standard Deviation} &= \sqrt{0.0123 \text{ mm}} \\ &\approx 0.111 \text{ mm} \end{aligned}$$

The maximum error is 0.15 mm, and the standard deviation is 0.111 mm, the accuracy is $0.111 \text{ mm} - 0.15 \text{ mm} = -0.039 \text{ mm}$. The negative sign indicates that the physical model is slightly larger than the digital model on average, resulting in a negative accuracy value.

5 RESULTS AND ANALYSIS

Here digital, and physical models discussed all parameters in the previous section, and according to that all parameters and research studies have been used in further experimental models. Furthermore, in this study, a total of two cubes and three orthosis models have been experimented on, of which one cube and one orthosis detailed analysis and measurements have been done in comparison of each in both the digital and physical models.

5.1 PRESENTATION OF THE COMPARISON RESULTS

5.1.1 RIGHT-HAND ORTHOSIS

Here, will discuss hand orthosis which is used in the medical field. This was created using a 3D printer. As discussed, and experimented with a cube in the previous section, accordingly, an orthosis will be experimented with in this section, and a try will be made to maintain quality in overall production. The figure for a right-hand orthosis is shown below. It is made using PLA material.

5.1.2 COMPARISON OF THE DIGITAL MODEL

Here, the digital model will be compared with the digital model but that will be at different layers from each other, as we compared the cube in the previous section as well, but here is a comparison of the top view of the orthosis which you can see one by one in the figure given below.

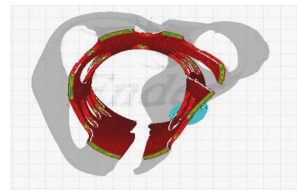


Figure 3: Layers at 902

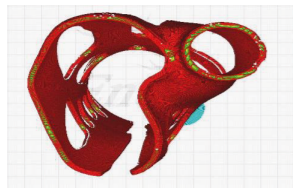


Figure 4: Layers at 1219

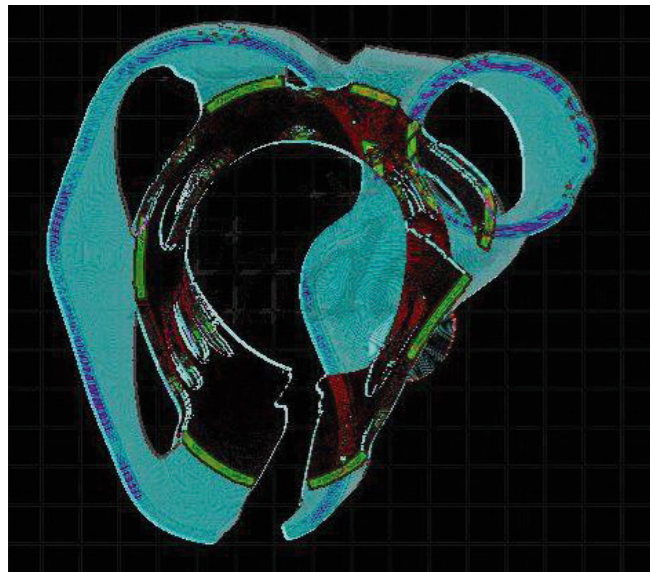


Figure 5: Results of the layers 902 and 1219

In the section, the figures above illustrate this, Figure 5 is the outcome of comparing Figures 3 and 4 in Python, while Figure 4 is Layers 1219. Here, we will discuss the comparison of the digital model layers results from the figure. But, in these figures, different colors are mentioned the green color indicates the surface of the first layer means the 902nd layer of the model, the purple color indicates the surface of the last layer means the 1219th layer and the sky-blue color indicates a different between both layers. In summary, red color is seen in the figure which indicates the contrast of the color and not a difference but enhances the visual contrast in the image. In addition, one more color seen in this figure is white, it is indicated somehow Z seam drawback, and it could be improved by printing parameters. Errors have been found due to which this model cannot be considered perfectly accurate but appears to be the ideal shape. In the below figure, we can see the actual dimension length, width, and height of the Orthosis in Cura software.

5.1.3 COMPARISON OF THE DIGITAL MODEL AND PHYSICAL MODEL



Figure 6: Digital Model of Orthosis Figure 7: Physical Model of Orthosis



Figure 8: Results of the Digital and Physical Model of Orthosis

The figures above illustrate this. Figure 8 is the outcome of comparing Figures 6 and 7 in Python, while both figures are the Digital model of orthosis and the Physical model of orthosis. In figure indicates the different colors. This figure, which the white color indicates is a digital model, it was created by blender software and taken by Cura slicing software. the gray color indicates a printed model means the actual model, was printed by an FDM printer. The physical model slightly shifted from a digital model and created a contrasting color brown. the reason behind that shift, as we discussed in the previous part, the camera angle and distance from the object are different. The scale for the tolerance of this model is ± 0.5 mm (lower limit: ± 0.5 mm). The digital model value was 78.50 mm(length), printed model value according to deviation, ± 0.5 mm. the accuracy of the results can be seen in Table 4. The digital photography method has been used in comparison to the model. The mea-

surements were taken by vernier calipers for the physical model, and according to these measurements (Table 2), the physical model's accuracy and quality can be good.

5.2 ANALYSIS OF THE PRINTOUT

The analysis of the printout, outcomes demonstrated that the physical model was precise, with the physical and digital models only slightly differing. The differences fell within a reasonable tolerance range, proving that the 3D printer and printing method could successfully create a model that was very similar to the digital model. However, it was challenging to capture images of the digital model and the printed model from the same distance and viewpoint, which made it challenging throughout the comparison process.

The table below shows the dimensional correctness of the digital model and the physical model. Measurements taken using vernier calipers are frequently accurate to 0.02 mm (0.001 inches) or greater. This model's measurement tolerance has a scale of 0.5 mm (lowest limit: 0.5 mm). Vernier calipers were used for all measurement taking. The DIN 862 Caliper Accuracy Standards were used for these vernier measurements.

	CUBE	CUBE	ORTHOSIS	ORTHOSIS
	Digital	3D Printed	Digital	3D Printed
Length (mm)	25.00	25.07	78.50	78.59
Width (mm)	25.00	25.02	100.70	100.62
Height (mm)	25.00	24.97	245.90	245.75

Table 2: Dimension of the design model and print model

5.3 FUTURE PROSPECTS

This thesis work already accomplishes the thesis's purpose. The digital model and physical model were compared using Python and slicing software, and the results of this thesis demonstrate that this strategy is both more practical and less expensive than other approaches. However, there are certain shortcomings and more ideas. As per my research and opinion, Comparing the digital model to the real model, it appears to be challenging and tough to take pictures of both the digital and real models. because it can be a little challenging for a human to photograph both at the same perspective and distance from the figure. In any case, the approach

for comparison is quite effective, and since the model error can be reduced at each level, it can be followed up with more work in the future. In general, these study topics hold promise for enhancing the precision, effectiveness, and dependability of 3D printing, which could have important ramifications for a variety of sectors.

6 CONCLUSION

This thesis discusses the research study's exploration of numerous methodologies for quality evaluation of the correctness of a 3D-printed model, emphasizing the use of Digital Photographic Measurement as a cost-effective and straightforward method. This research used Python and Slicing software to create a quality control system for 3D-printed medical devices. The primary goal is to evaluate the 3D printing process by comparing the produced 3D model to the original QA phantom using image-based measures. The investigation shed important light on the difficulties and possibilities involved in judging the accuracy and caliber of 3D-printed models. Overall, the information gathered from the literature research represents a thorough grasp of where the area of 3D printing and quality assurance is at the moment. Furthermore, it is believed that this study's findings would contribute to the expansion and improvement of 3D printing

technology by guiding the direction of future research and development efforts.

The present work suggests that manual pictures can be upgraded with some new techniques. A direct comparison between the two models, however, proved difficult because of variations in camera angle and distance. Python was unable to produce correct comparison results because of the misalignment, which led to disparities. The limits brought on by camera angle and distance variations have a considerable effect on the alignment procedure. It might be challenging to make an accurate comparison because of the distortions and misalignments that these differences can cause between the digital and physical representations. It is crucial to be aware of this study's limitations. The primary areas of study include size, shape, and surface polish. Nevertheless, despite these drawbacks, this research offers insightful information about the potential of digital models for quality control reasons.

In summary, for the goal of quality assurance in 3D printing, this project proved the value of comparing digital models with production in models. The object quality of the 3D-printed objects was assessed using slicing soft-

ware and Python, revealing areas for improvement, and offering insightful information for future developments. Overall, the study helps improve the quality of the 3D printing process and object.

REFERENCES

[1] Carmel, "Sculpteo," Die Geschichte des 3D-Drucks, 26 April 2022. [Online]. Available: <https://www.sculpteo.com/de/3dlerzentrum/die-grundlagen-des-3d-druckens/die-geschichte-des-3d-drucks/>. [Accessed 28 April 2023].

[2] R. Omari, C. Hunt, J. Coumbaros and B. Chapman, "Virtual anthropology? Reliability of three-dimensional photogrammetry as a forensic anthropology measurement and documentation technique.," International Journal of Legal Medicine, vol. 135, pp. 939-950, 2021.

[3] S. Nuchitprasitchai, R. Michael and M. P. Joshua, "Factors affecting real-time optical monitoring of fused filament 3D printing.," Progress in Additive Manufacturing, vol. 2, pp. 133-149, 2017.

[4] Z. Jin, Z. Zhizhou and X. G. Grace, "Autonomous in-situ correction of fused deposition modeling printers using computer vision and deep learning.," Manufacturing Letters, vol. 22, pp. 11-15, 2019.

[5] M. Moretti, R. A. and S. N., "In-process simulation of the extrusion to support optimization and real-time monitoring in fused filament fabrication.," Additive Manufacturing, vol. 38, p. 101817, 2021.

[6] B. Orzeł and S. Kinga, "Comparison of 3D Printout Quality from FDM and MSLA Technology in Unit Production.," Symmetry 14, vol. no. 5, p. 910, 2022.

[7] M. Pańnikowska-Łukaszuk, K. Katarzyna, K. Kamil, U. Arkadiusz and W.-C. Magda, "Time distribution analysis of 3D prints with the use of a filament and masked stereolithography resin 3D printer.," Advances in Science and Technology. Research Journal 16, vol. no. 5, pp. 242-249, 2022.

[8] V. Uralde, V. Fernando, A. Eider, S. Alfredo, and B. Tomas, "Symmetry and Its Application in Metal Additive Manufacturing (MAM).," Symmetry 14, vol. no. 9, p. 1810, 2022.

[9] J. Qiu, H. Kun, A. D. Brandon, C. Jyh-Cheng, S. Liting, S. Yong and X. Longchun, "Constructing customized multimodal phantoms through 3D printing: A preliminary evaluation.," Frontiers in Physics, vol. 9, p. 605630, 2021.

[10] J. L. Martinez, M. Mariano, M. Jesus, R. Alfredo, and S. Manuel, "Supervised learning of natural-terrain traversability with synthetic 3D laser scans.," Applied Sciences 10, vol. no. 3, p. 1140, 2020.

[11] K. Rajan, S. Mahendran, K. Kumaran, W. H. Wan Sharuzi and R. Md Mustafizur, "Fused deposition modeling: Process, materials, parameters, properties, and applications.," The International Journal of Advanced Manufacturing Technology 120, Vols. no. 3-4, pp. 1531-1570, 2022.

[12] M. Jani, "All3DP," 12 March 2023. [Online]. Available: <https://all3dp.com/1/cura-tutorial-software-slicer-cura-3d/>. [Accessed 11 May 2023].

[13] S. Habib, "PlasticRanger," 09 February 2023. [Online]. Available: <https://plasticranger.com/density-of-abs-material/>. [Accessed 11 May 2023].

[14] A. AYDIN, F. Kahraman and I. Vjekoslav, "CAN HUMAN BEINGS PRODUCE EVERYTHING? THE LIMITS OF THE 3D PRINTERS.," RESEARCH & REVIEWS IN ENGINEERING-Summer, 2019, p. 137, 2019.

Vergleich von optischen 3D-Sensoren zur Feinreferenzierung eines Klebprozesses für schiffbauliche Komponenten

Till Braun, Christoph Schack, Wilko Flügge

Fraunhofer-Institut für Großstrukturen in der Produktionstechnik IGP,
Albert-Einstein-Straße 30, D-18059 Rostock
E-Mail: till.braun@igp.fraunhofer.de
<http://www.igp.fraunhofer.de>

Abstract. Im Bereich der industriellen Bildverarbeitung haben optisch messende 3D-Sensoren für Handhabungs- und Montageaufgaben in den letzten Jahren zunehmend an Bedeutung gewonnen. Im Folgenden wird ein geometrisches Referenzierungsverfahren für einen mobilen Klebroboter vorgestellt. Das Verfahren zur Erfassung der Fügestelle basiert auf der Open-Source Bibliothek Open3D. Dafür werden die zwei Messprinzipien Stereovision und Musterprojektion verwendet. Die Robustheit der ausgewählten optischen 3D-Sensoren wird für das entwickelte Verfahren experimentell untersucht und verglichen. Im Vordergrund steht die Überprüfung der Eignung unter schiffbaulichen Umgebungsbedingungen. Dazu wurden mehrere Messreihen mit reflektierenden Bauteilen durchgeführt, der Störeinfluss von Fremdlicht beurteilt und ein geeigneter Matching-Algorithmus für die Stereokamera anhand einer Clusteranalyse der Punktwolke ausgewählt. Die Stereokamera lieferte unter Fremdluchteinfluss die besten Ergebnisse. Bei der Clusteranalyse ergab sich ein maximales Rauschen von $r_{np} = 0,036\%$. Das Musterprojektionssystem lieferte ein Rauschsignal von $r_{np} = 4,13\%$.

1 EINLEITUNG

Im Gegensatz zu dem Schweißen ist das Fügeverfahren des Klebens im Schiffbau bisher wenig automatisiert und kommt anders als im Bootsbau, nur selten für strukturelle Verbindungen zu Einsatz. Im Ausrüstungsbereich wird es für spezielle Anwendungen eingesetzt wie die Montage von Halterungen oder das Verkleben von Scheiben [1]. Das Kleben bietet dabei viele Vorteile. Bspw. ermöglicht es das Fügen von unterschiedlichen Materialien, die anders nicht zuverlässig verbunden werden können. Weitere wünschenswerte Eigenschaften liegen in der Vibrationsdämpfung und dem Vermeiden von Korrosionsproblemen, die bei Fügeverfahren mit metallischen Verbindungselementen auftreten [1].

Um das Kleben im Schiffbau zu etablieren und Zulassungen durch Klassifikationsgesellschaften wie den DNV zu erreichen, bedarf es eines wiederholgenauen, qualitätsgesicherten Prozesses sowie der Durchführung von Tests zur Beurteilung der Dauerfestigkeit der hergestellten Klebverbindungen [1]. Für die Dokumentation relevanter Prozessdaten der Qualitätssicherung soll das Verfahren daher automatisiert und Umgebungsdaten, die Einfluss auf den Klebprozess haben (bspw. die Abbindebedingungen [2]), sensorisch erfasst werden.

Der Klebprozess unterteilt sich in die Vorbehandlung, die Klebstoffapplikation sowie das Fügen der Bauteile. Der gesamte Prozess soll zunächst unter der Aufsicht eines Bedieners halb automatisch ablaufen. Neben der Überwachung des Roboters wird der Bediener in den Qualitätssicherungsprozess einbezogen, um bspw. eine Grobsichtprüfung der Fügestelle durchzuführen und etwaige Kratzer oder Verfärbungen der Beschichtung

oder Oberflächenverschmutzung zu detektieren, die die Qualität der Klebverbindung beeinflussen. Basierend auf den Benutzereingaben müssen ggf. einzelne Schritte der Vorbehandlung wie das Reinigen der Oberfläche durch den Roboter wiederholt werden.

Zur Erzeugung von Bahnbefehlen für die Vorbehandlung und den anschließenden Klebstoffauftrag auf dem Bauteil durch den Knickarmroboter muss die exakte Pose der mobilen Plattform relativ zu der Fügestelle bestimmt werden. Für die geometrische Referenzierung wird nach der Vorpositionierung eine Messfahrt durchgeführt, bei der die Fügestelle großflächig erfasst und Toleranzen ermittelt werden.

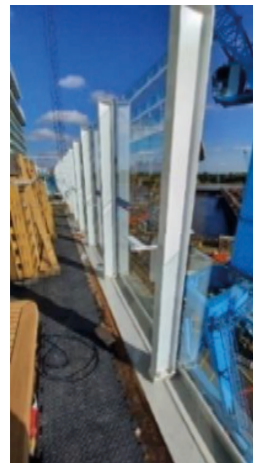


Abbildung 1: Halter einer Windschutzscheibe

Ziel der Referenzierung ist daher neben der Definition des Bauteilkoordinatensystems das Ermitteln von Korrekturwerten für die Bahn des Roboters aus den Messwerten.

Neben Sensoren zum Sammeln von Umgebungsdaten wie Temperatur und Luftfeuchtigkeit zur Qualitätssicherung werden für die Automatisierung des Klebprozesses geeignete Sensoren zur geometrischen Erfassung des Bauteils benötigt, die unter Störeinflüssen wie künstlichen Lichtquellen oder reflektierenden Oberflächen verlässliche Messwerte liefern.

In den nachfolgenden Versuchen soll daher erprobt werden, welche flächig erfassende 3D-Sensoren sich für die Referenzierung schiffbaulicher Teile eignen und welche Sätze von Parametern zu verwenden sind. Weiterhin soll das entwickelte Verfahren zur benutzergeführten Programmierung des Roboters an realen Bauteilen aus unterschiedlichen Materialien bzw. reflektierenden Oberflächen sowie unter dem Störeinflussfaktor des künstlichen Lichts getestet werden.

2 METHODIK UND VORGEHENSWEISE

2.1 SENSOREN UND MATCHING-ALGORITHMEN ZUR BAUTEILERFASSUNG

Zur optischen Erfassung von geometrischen Messdaten können die drei Messverfahren des Laufzeit-,

interferometrischen- und Triangulationsverfahren unterschieden werden. Weiterhin kann die Erfassungsart der Geometrie in punktförmig, linienförmig und flächig erfassende Verfahren klassifiziert werden [3].

Zu den flächig erfassenden Verfahren gehören u.a. die in dieser Arbeit betrachteten Messverfahren der Strukturlichtverfahren sowie die passive und aktive Stereoskopie. Weitere wichtige Sensoren, die u.a. auch in der mobilen Robotik zur Umgebungswahrnehmung eingesetzt werden, sind Time-of-Flight-Kameras sowie laufzeitbasierte oder durch Phasenkorrelation messende Mehrlängenscanner [4].

Die Erweiterung des Lichtschnittverfahrens durch die Projektion mehrerer Linien in einer Ebene erlaubt die flächige Erfassung eines Bauteils mithilfe von strukturierten Lichtmustern. Analog zu den linienförmig erfassenden Lichtschnittsensoren erfolgt die Gewinnung von Tiefeninformationen mit dem Triangulationsprinzip durch die Beobachtung der Reflektion der Laserlinien mithilfe einer Matrixkamera [5]. Anstatt einzelner Laserlinien werden strukturierte Lichtmuster auf das Bauteil projiziert, die in den Kamerabildern wiedererkannt werden. Verwendete Muster können Schachbrett- [6], Lichtpunktfeld- oder Lichtgittermuster [5] sein. Die eindeutige Identifikation der einzelnen Liniennummern und somit auch der Tiefeninformationen kann durch eine Kodierung der Fläche, Tiefe oder Zeit erreicht werden [6]. Dies ist notwendig, da Stufen oder Sprünge auf der Bauteiloberfläche die einzelnen projizierten Laserlinien unterbrechen und somit die Zuordnung verhindern können [6].

Bei der passiven Stereoskopie wird eine Szene durch mehrere einzelne Aufnahmen einer Kamera oder durch die Kombination mehrerer Kamerasysteme aus unterschiedlichen Ansichten erfasst [6]. Die Gewinnung von Tiefeninformationen erfolgt durch die Suche korrespondierender Punkte in beiden Bildern. Ein gängiger Matching-Algorithmus zum Lösen des Korrespondenzproblems ist das Semi-Global Matching (SGM) [7]. Bei diesem Verfahren wird eine globale Kostenfunktion durch die pixelweise Berechnung und Akkumulation von Matchingkosten in horizontaler, vertikaler sowie diagonaler Richtung approximiert. Hierbei erfolgt eine Verknüpfung mit einem Glattheitskriterium [8]. Das SGM gilt als robustes Verfahren gegenüber kleineren Spiegelungen, da die pixelweise Berechnung der Matchingkosten durch Transinformation, der Stärke des statistischen Zusammenhangs der beiden

Bilder, berechnet wird [8].

Das Korrespondenzproblem wird bei der aktiven Stereografie durch das Projizieren eines Lichtmusters auf das Bauteil gelöst. Die Suche des Korrespondenzpunkts wird auf die Suche einzelner Lichtpunkte in beiden Bildern der Stereokamera begrenzt. Durch die Verwendung komplexerer Muster kann die Anzahl der Korrespondenzen pro aufgenommenen Bildpaar gesteigert werden. Durch die zeitliche Kodierung bspw. nach dem Grey-Code Verfahren kann auch die Datenrate des Sensors verbessert werden [6].

2.2 VERWENDETE SOFTWARE

Für die Visualisierung und Verarbeitung von Punktwolken existiert u.a. die Open-Source-Bibliothek Open3D [9], die Algorithmen zur Registrierung und Verarbeitung von Punktwolken in Python und C++ beinhaltet. In dieser Arbeit wurde die Python-Implementierung von Open3D verwendet, da die Programmierschnittstellen (API) der getesteten Sensoren ebenfalls Python-Implementierungen beinhalten. Im Fall der Photoneo MotionCam-3D M+ (nachfolgend Photoneo M+ genannt) ist eine Python-Implementierung des GigE Vision Standards unter [10] sowie bei der Ensenso C57m ein Python Interface des EnsensoSDK [11] verfügbar. Die erprobten Matching-Algorithmen der Stereokamera sind ebenfalls Bestandteil von [11]. Ein weiterer Vorteil der Python-Implementierung von Open3D ist die Kompatibilität mit der Bibliothek NumPy bspw. für den Import von Messdaten, die grundlegende optimierte Funktionen für wissenschaftlichen Rechenoperationen enthält [12].

2.3 VERSUCHSAUFBAU

Um eine robuste Bauteilerfassung zu gewährleisten, soll ein flächig erfassender, optischer Sensor mit einer Genauigkeit und Auflösung von unter einem Millimeter bei einem Abstand von anderthalb Metern ausgewählt werden. Tabelle 1 vergleicht die grundlegenden Daten der ausgewählten aktiven Stereographie-Kamera Ensenso C57m sowie dem am Fraunhofer IGP verfügbaren Musterprojektionssystem Photoneo M+.

	Photoneo M+	Ensenso C57m
Größe	80 x 68 x 427 mm	506x 58x 100 mm
Auflösung	1120 x 800	2472 x 2064
Genauigkeit	0,6 mm bei z = 1574 mm	0,3 mm bei z = 1600 mm
Bilder pro Sekunde	20	3
Arbeitsabstand	630 – 1574 mm	650 - 5000 mm

Tabelle 1: Auszug der technischen Daten der getesteten Sensoren [14], [15]

Für die Erprobung der beiden Sensorsysteme werden zwei Mikropaneele aus Aluminium und Stahl verwendet. Der Versuchsaufbau ist in Abbildung 2 dargestellt. Beide Sensoren werden durch einen an einem Gantry montiertem Kuka KR16-1610 Roboter positioniert. Nach der Montage des Sensors wird zur Bestimmung der Transformationsmatrix zwischen dem Flansch- und dem Sensorkoordinatensystem zunächst eine Hand-Augen-Kalibrierung durchgeführt. Um den Einfluss von Fremdlicht zu testen wurde der in Abbildung 2 dargestellte Scheinwerfer verwendet.

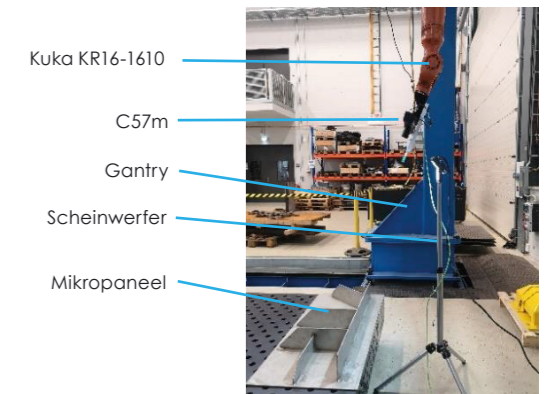


Abbildung 2: Versuchsaufbau C57m und Stahl Mikropaneel

Die Beleuchtungsstärken der in Tabelle 2 aufgeführten Szenarien wurden mittig auf dem Paneel mit einem PeakTech P 5086 Lux-Messgerät [15] gemessen.

	Tageslicht	Tageslicht mit Fremdlicht	Abgedunkelt mit Fremdlicht	Abgedunkelt
Beleuchtungsstärke	425 lx	1000 lx	570 lx	0,6 lx

Tabelle 2: Lichtverhältnisse bei Messreihen

2.4 VERSUCHSDURCHFÜHRUNG

Im ersten Schritt wurde für das Szenario „Tageslicht mit Fremdlicht“ eine Messreihe von Einzelaufnahmen zur Auswahl eines geeigneten Matching-Algorithmus der Stereographie-Kamera durchgeführt. Für die Aufnahmen wurde der Projektor eingeschaltet und je 32 Bildpaare bzw. Musterverschiebungen für das Matching verwendet. Die gewählte Belichtungszeit wurde mit einem Wert von ca. 1,3 ms automatisch ermittelt. Im nächsten Schritt wurden die Messdaten vorgefiltert, um doppelte Mess- sowie Nullwerte zu entfernen und das Rauschen zu reduzieren. Für die statistische Entfernung der Ausreißer wurden zunächst für jeden Punkt pi die drei nächsten Nachbarn mithilfe eines k-d-Baums bestimmt. Als Outlier werden anschließend die Punkte gefiltert, die einen Abstand von $3\sigma \leq d$ zu ihren Nachbarn aufweisen. Als Schwellwert wurde entsprechend der geforderten Auflösung $d = 1$ mm gewählt. Weiterhin wurden Messwerte außerhalb des Arbeitsbereichs des Sensors gefiltert.

Zur Auswahl eines Matching-Algorithmus sowie zur Beurteilung des Störeinflusses des Fremdlichts wurde eine Clusteranalyse der Messdaten vorgenommen. Dabei wurde der in der Bibliothek *sklearn* [16] enthaltene DBSCAN-Algorithmus (Density-Based Spatial Clustering of Applications nach [17]) verwendet. Als Parameter der maximalen Distanz zwischen zwei Samplepunkten wurde $\epsilon = 3$ mm sowie ein Minimum von 10 Nachbarpunkten zu einem Kernpunkt gewählt. Als Bewertungskriterien wurden die Anzahl der Messpunkte n_p , Anzahl der Cluster n_c , die durchschnittliche Anzahl der Punkte pro Cluster p_c , das Verhältnis der Rauschpunkte des DBSCAN-Algorithmus zu den Messpunkten r_{np} , sowie das Verhältnis der statistisch gefilterten Punkte zu den Messpunkten r_{fp} bewertet.

3 ERGEBNISSE

3.1 EINFLUSS DER MATCHING-ALGORITHMEN

Die Ergebnisse der Einzelaufnahmen mit unterschiedlichen Matching-Algorithmen sind in Tabelle 3 aufgeführt. Durch die höhere Anzahl der erfassten Messpunkte bei dem „PatchMatch“ Verfahrens erhöht sich die Dichte der Punktwolke, was zu einer vergleichsweise geringeren

Matching-Algorithmus	n_p	n_c	p_c	r_{np}	r_{fp}
PatchMatch	3.493.991	373	9.349	0,196 %	0,006 %
SGM Aligned and diagonal paths	940.511	632	1.485	0,234 %	4,2 %
SGM Axis Aligned Paths	909.597	438	2.073	0,146 %	3,1 %
SGM Diagonal Paths	926.924	627	1.473	0,315 %	4,9 %
SGM Single Pass	1.002.862	410	2.441	0,181 %	4,2 %
Sequence Correlation	562.721	440	1.279	0,031 %	4,2 %

Tabelle 3: Ergebnisse Clusteranalyse Matching-Algorithmen C57m

Anzahl von Clustern und einem geringeren Rauschverhältnis r_{fp} führt.

Der Einfluss der unterschiedlichen Matching-Algorithmen ist ebenfalls in den Disparity Maps in Abbildung 3 zu beobachten. Während bei der Sequence Correlation und dem SGM nur vereinzelte Punkte auf dem Bauteil erfasst wurden (3b – 3f), erreicht das PatchMatch-Verfahren die größte Abdeckung von Messpunkten auf dem Paneel.

3.2 EINFLUSS DER BELEUCHTUNG

Um den Einfluss von Fremdlicht zu prüfen, wurde das Aluminiumpaneel unter den Belichtungsszenarien aus Tabelle 1 und dem Parametersatz „Static Scenes Quality“ [18] mit der Stereographie-Kamera aufgenommen. In den Ergebnissen aus Tabelle 4 lässt sich ein geringer Einfluss des abgedunkelten Labors feststellen, der sich in einem niedrigeren Verhältnis der Rauschpunkte sowie einer erhöhten Anzahl von Clusterpunkten bei einer gleichzeitig niedrigeren Anzahl zeigt.

Belichtungsszenario	n_p	n_c	p_c	r_{np}	r_{fp}
Tageslicht	670.321	73	9.179	0,036 %	4,7 %
Tageslicht mit Fremdlicht	660.615	57	11.563	0,222 %	3,5 %
Abgedunkelt mit Fremdlicht	662.047	49	13.484	0,198 %	3,5 %
Abgedunkelt	653.951	32	20.432	0,016 %	3,5 %

Tabelle 4: Ergebnisse Clusteranalyse Lichteinfluss C57m

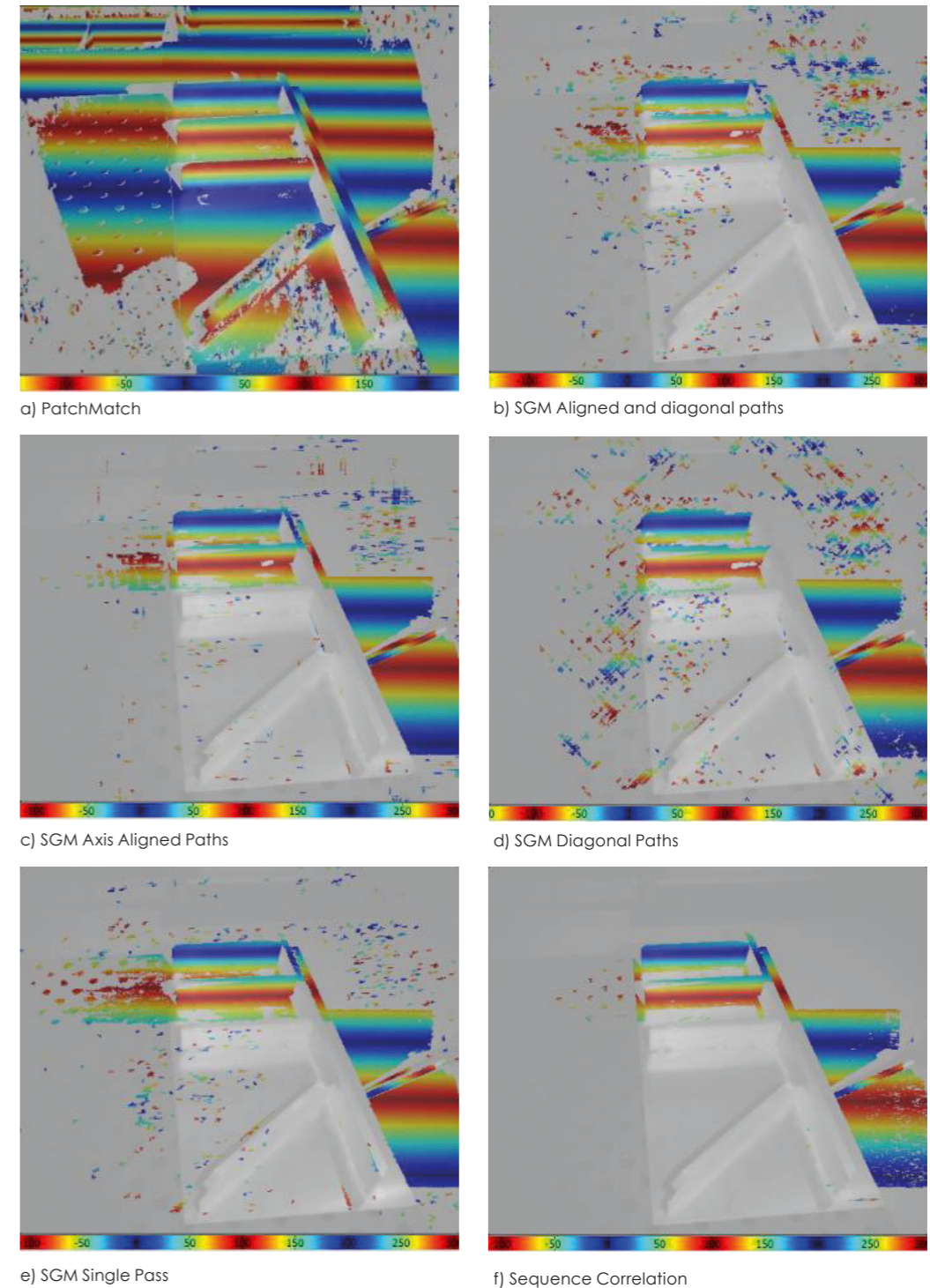


Abbildung 3: Einfluss der Matchingstrategien in Tiefenbild

Die Vorgehensweise wurde entsprechend für den Musterprojektionssensor Photoneo M+ wiederholt. Die Messwerte aus Tabelle 5 wurden im „Dynamic Mode“ des Sensors (siehe [13]) und mit einer Belichtungszeit von 10,24 ms aufgenommen. Auch hier lässt sich eine ähnliche Tendenz bei der Abdunklung des Raums beobachten.

Belichtungsszenario	n_p	n_c	p_c	r_{np}	r_{fp}
Tageslicht	431.399	651	636	4,02 %	2,2 %
Tageslicht mit Fremdlicht	431.370	688	601	4,13 %	4,4 %
Abgedunkelt mit Fremdlicht	431.672	608	685	3,51 %	2,6 %
Abgedunkelt	425.572	572	720	3,22 %	2,89 %

Tabelle 5: Ergebnisse Clusteranalyse Lichteinfluss Photoneo M+

4 ENTWICKLUNG EINES BENUTZERGEFÜHRTEN REFERENZIERUNGSVERFAHRENS

Unter der Anforderung eines halbautomatisierten Klebprozesses und den Eingaben eines Benutzers auf einem Tablet-PC wurde basierend auf der Visualization GUI von Open3D die grafische Benutzeroberfläche aus Abbildung 4 entwickelt, die die manuelle Definition des Bauteilkoordinatensystems der Fügestelle erlaubt.

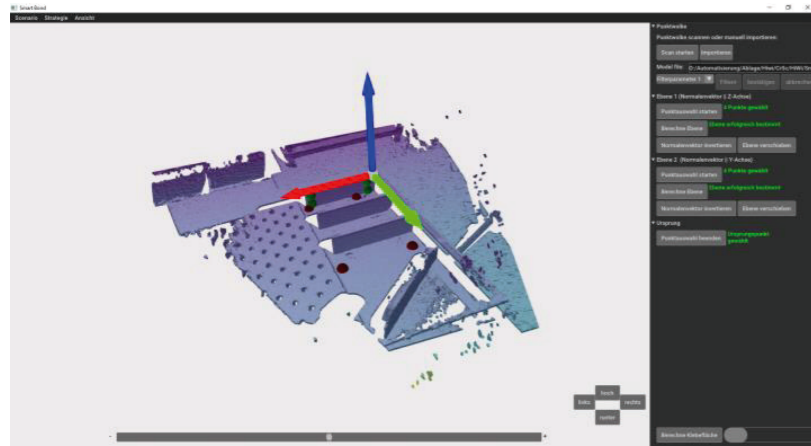


Abbildung 4: Benutzeroberfläche Definition Bauteilkoordinatensystem

Nach der Grobpositionierung der mobilen Plattform und der Durchführung der Messfahrt wird die resultierende Punktwolke zunächst, wie in dem letzten Kapitel beschrieben, gefiltert. Im nächsten Schritt wird die bereinigte Punktwolke durch Voxel diskretisiert und in der GUI visualisiert. Zum Setzen des Bauteilkoordinatensystems stehen dem Benutzer die Strategien der Auswahl von drei Ebenen oder alternativ zwei Ebenen und einem Ursprungspunkt zur Verfügung. Nach der Auswahl von mindestens vier Punkten pro Ebene werden umliegende Messpunkte in einem rechteckigen Volumen ausgeschnitten. Aus den Hilfspunkten wird mithilfe eines RANSAC Algorithmus (Random Sample Consensus [19]) die Ebenenparameter sowie der resultierende Normalenvektor bestimmt. Aus den Normalenvektoren der Ebene und ggf. der Wahl des Ursprungspunkts ergeben sich Lage und Orientierung des Bauteilkoordinatensystems. Der beschriebene Ablauf wurde mit den in den Versuchen aufgenommenen Punktwolken erprobt. Beide getesteten Systeme sind in der Lage, ausreichend robuste und detaillierte Aufnahmen zu liefern.

5 DISKUSSION UND AUSBLICK

Basierend auf der Bibliothek Open3D wurde eine grafische Benutzeroberfläche zur manuellen Bauteilreferenzierung auf einem Tablet entwickelt und mit den aufgenommenen Sensordaten getestet. Der Versuch zur Auswahl eines geeigneten Matching-Algorithmus zeigt, dass die Wahl des Matching-Algorithmus sensible Auswirkungen auf das Ergebnis der Punktwolke der Stereokamera hat. Das SGM erwies sich für die getesteten spiegelnden Bauteile als unpraktikabel, weshalb das Pattern Matching Verfahren ausgewählt wurde. Der Einfluss von indirekten Fremdlicht war in dem durchgeführten Versuch bei beiden Sensoren vernachlässigbar gering. Beide Messverfahren lieferten nach der Vorverarbeitung der Messdaten in Bezug auf die Anforderung der Auflösung ausreichend gute Ergebnisse. Die Referenzierung konnte mit den Messdaten beider Sensoren, den verwendeten Bauteilen und dem entwickelten Verfahren durchgeführt und ein Bauteilkoordinatensystem definiert werden.

Eine praktische Validierung der erreichbaren Lagetoleranz des Bauteilkoordinatensystems konnte innerhalb des Leihzeitraums der Stereokamera nicht durchgeführt werden, weshalb die Automatisierung des Klebprozesses nur eingeschränkt zu beurteilen ist. Die erreichbare Lagetoleranz kann bspw. mithilfe eines Lasertrackers ermittelt werden, indem durch den Roboter nach der Definition des Bauteilkoordinatensystems mehrere Bahnpunkte angefahren, markiert und anschließend die durchschnittliche Abweichung ausgewertet wird. Der nächste Schritt besteht in der Implementierung geeigneter Methoden zur Erfassung fertigungsbedingter Toleranzen mit dem Ziel der Bahnkorrektur des Roboters.

LITERATURVERZEICHNIS

- [1] Center of maritime Technologies gGmbH (CMT): Untersuchungen zum aktuellen internationalen Stand der Technik bezüglich Entwicklung und Anwendung von Klebeverbindungen in der Schiffsfertigung; Studie; [online] [Zugriff am: 10.09.2023]; Verfügbar unter: <https://dmz-maritim.de/wp-content/uploads/2022/05/Studie-Klebeteknologie-Schiffsfertigung.pdf>; 2020 – 05
- [2] Rasche, M.: Qualitätsbestimmende Einflussgrößen bei Kunststoff-Metall-Klebeverbindungen; Schweißtechnische Forschungsberichte; Band 5; Deutscher Verlag für Schweißtechnik (DVS); ISBN 3-87155-864-8; Verfügbar unter: doi.org/10.25968/opus-38; Düsseldorf; 1986
- [3] Berndt, D.: Flexibel und adaptierbar: Stand der Technik und Zukunft optischer 3D-Messtechnik; Qualität und Zuverlässigkeit : QZ; Volume 58; ISSN 0033-5126; Carl Hanser Verlag GmbH & Co. KG; S. 98 – 101; 2013
- [4] Weber, H.: Funktionsweise und Varianten von LiDAR-Sensoren; Whitepaper | Sick; Volume: 8022039; [online] [Zugriff am: 03.10.2023]; Verfüg-

- bar unter: https://cdn.sick.com/media/docs/5/25/425/whitepaper_lidar_de_im0079425.pdf; 2018 – 07
- [5] Hesse, S., Schnell, G.: Sensoren für die Prozess- und Fabrikautomation; 6. Auflage; Springer Vieweg Wiesbaden; ISBN 978-3-658-05867-8; Verfügbar unter: doi.org/10.1007/978-3-658-05867-8; 2014
- [6] Hamfeld, H.: Aktive Stereoskopie, Neue Verfahren zur dreidimensionalen Vermessung von Objekten; Dissertation; Kaiserslautern; 09 – 2001
- [7] Hirschmüller, H.: Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information; 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05); Verfügbar unter: doi.org/10.1109/CVPR.2005.56; San Diego; 2005; S. 807 – 814;
- [8] Föll, G.: Stereo Vision: Vergleich verschiedener Algorithmen zur Lösung des Korrespondenzproblems; Ausarbeitung; Fakultät Technik und Informatik; Hochschule für Angewandte Wissenschaften Hamburg; 2010 – 09
- [9] Zhou, Q., Park, J., Koltun, V.: Open3D: A Modern Library for 3D Data Processing; Verfügbar unter: doi.org/10.48550/arXiv.1801.09847; 08 – 2018
- [10] Photoneo s.r.o.: Photoneo Python examples; [online] [Zugriff am: 14.10.2023]; Verfügbar unter: <https://github.com/photoneo-3d/photoneo-python-examples>
- [11] Optonic GmbH: Python Interface; [online] [Zugriff am: 14.10.2023]; Verfügbar unter: <https://www.ensenso.com/manual/3.5/python/index.html>
- [12] Harris, C.R., Millman, K.J., van der Walt, S.J. et al.: Array programming with NumPy; Nature; Volume 585; Verfügbar unter: doi.org/10.1038/s41586-020-2649-2; 2020 – 09
- [13] Photoneo s.r.o.: MotionCam-3D M+; [online] [Zugriff am: 17.10.2023]; Verfügbar unter: <https://www.photoneo.com/de/products/motioncam-3d-m-plus/>
- [14] IDS Imaging Development Systems GmbH: C57-6-M - Spezifikation; Datenblatt; Obersulm; 03 - 2023
- [15] PeakTech Prüf- und Messtechnik GmbH: «PeakTech® P 5086» Digitales LED Lux-Messgerät, 3 3/4-stellig; Datenblatt; [online] [Zugriff am: 17.10.2023]; Verfügbar unter: <https://www.peaktech.de/datasheet/cfbefe5799f447ba9c758b5cd43fb98b>
- [16] Pedregosa, F. et al.: Scikit-learn: Machine Learning in Python; The Journal of Machine Learning Research; Volume 12; Verfügbar unter: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>; 2011 – 11; S. 2825 – 2830
- [17] Ester, M. et al.: A density-based algorithm for discovering clusters in large spatial databases with noise; KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining; AAAI Press; Verfügbar unter: <https://cdn.aaai.org/KDD/1996/KDD96-037.pdf>; 1996 – 08; S. 226 – 231
- [18] Optonic GmbH: Parameter Adjustment in NxView; [online] [Zugriff am: 17.10.2023]; Verfügbar unter: <https://www.ensenso.com/manual/3.5/about/nxview.html?highlight=static%20quality>
- [19] Fischler, M. A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography; Communications of the ACM; Volume 24; Verfügbar unter: doi.org/10.1145/358669.358692; 1981 – 06; S. 381 - 395

Virtuelles Fitting zur Unterstützung bei der Entwicklung von neuartigen Kanülen zur mechanischen Kreislaufunterstützung von Patienten mit univentrikulären Herzen

Brockmann H, Teßmann R, Menon AK

Berlin Heart GmbH,
Wiesenweg 10, 12247 Berlin
www.berlinheart.de

Abstract. Eine neuartige Y-förmige venöse Kanüle wird entwickelt, um das Produktportfolio des EXCOR-VAD (Berlin Heart GmbH) zu erweitern. Die Kanüle wird für die Versorgung von Patienten mit sehr spezieller und individueller Anatomie eingesetzt. Um die anatomischen Besonderheiten bei der Produktentwicklung berücksichtigen zu können, werden virtuelle Fittings mit verschiedenen Entwürfen der Kanüle mit Herzchirurgen und Entwicklern durchgeführt. Anpassungen können so gleich im Gespräch visualisiert und analysiert werden.

WISSENSCHAFTLICHER HINTERGRUND UND MOTIVATION

Die Kanüle soll speziell die Versorgung von Patienten mit komplexen, zentral-venösen Gefäßanomalien, z.B. von Fontan- oder Glenn-Patienten, gewährleisten. Glenn- sowie Fontan-Patienten sind mit einer Single-Ventrikel-Physiologie geboren und durch mehrere komplexe Korrektur-Operationen im frühen Kindesalter auf eine partielle bzw. komplette Fontan-Zirkulation umgestellt worden. Wenn diese Kreislauf-Konstellation, also die passive Perfusion der Lungenstrombahn) versagt, spricht man vom Failing-Fontan (sub-pulmonaler Typ). Bisher sind die therapeutischen Möglichkeiten nach dem Versagen der Fontan-Zirkulation gering. Die Kanüle in Kombination mit einem Herzunterstützungssystem soll für diese Herausforderungen eine neue, innovative Therapiemöglichkeit darstellen und so die Wartezeit bis zu einem Spenderorgan überbrücken.

VERWENDETE METHODEN

Aus insgesamt 26 Thorax MRT Aufnahmen von Fontan-Patienten werden, nach einer Vorauswahl, acht Datensätze für das virtuelle Fitting aufbereitet. Um eine 3D-Darstellung aus den Digital Imaging and Communications in Medicine (DICOM-) Daten zu erzeugen, werden die Datensätze mittels der kommerziellen Software Amira 2019.03 umgewandelt. Amira ist eine Software für die Visualisierung, Verarbeitung und Analyse von 3D und 4D Datensätzen. Mit dieser Software werden die einzelnen anatomischen Strukturen mittels der Festlegung von Schwellwerten und des Region Growing Algorithmus abgegrenzt. Folgende anatomische Strukturen werden bei jedem Datensatz segmentiert:

TCPC-Verbindung (totale cavopulmonale Anastomose), Herz, Lunge, Sternum, Rippen, Haut und Zwerchfell und danach als stl-Files für das CAD-Programm exportiert.

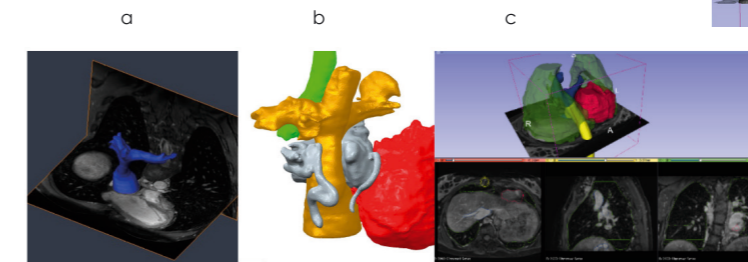


Abb. 1: Darstellung der DICOM-Ansicht mit segmentierter TCPC Verbindung im Programm Amira (a), Darstellung der anatomischen Strukturen in CAD (CreoParametric) (b), Software 3D Slicer mit DICOM und stl-Files zum Virtuellen Fitting der Kanülen (c)

Für das virtuelle Fitting der Kanülen werden Kanülen-Modelle als parametrisiertes Modell im CAD-Programm (Creo Parametric) erstellt. Hierbei ist es möglich, die Kanülenform an gegebene anatomische Besonderheiten anzupassen. In diesem Fall wurden folgende geometrischen Ausprägungen der Kanüle parametrisiert:

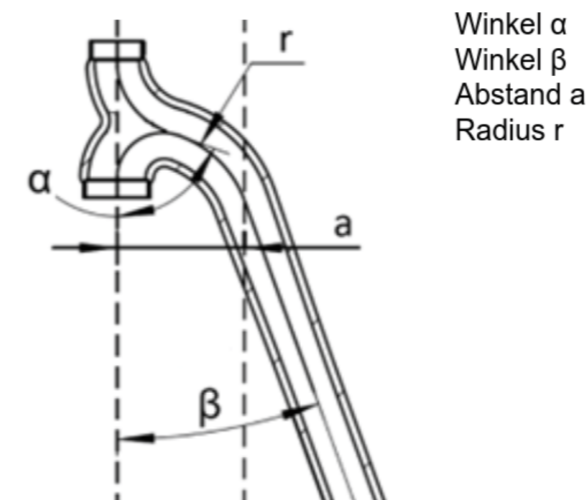


Abb. 2: Parametrisierbare Ausprägungen der Kanüle

Um die Fertigbarkeit der Kanüle als einfaches Spritzgussteil zu gewährleisten, werden geometrische Anpassungen nur in der xy-Ebene vorgenommen, um diese als Trennebene der Gussform zu erhalten.

Zur Bewertung der Kanülen werden neben der geometrischen Passung bei der Kanülenauswahl auch das Strömungsverhalten und die Fertigbarkeit berücksichtigt. Mithilfe des Programms 3D-Slicer werden nun die patientenspezifischen anatomischen Strukturen, sowohl als DICOM, als auch als 3D-Ansicht überlagert und es können die vorausgewählten Varianten der Kanülen eingblendet und deren Passgenauigkeit analysiert und diskutiert werden.

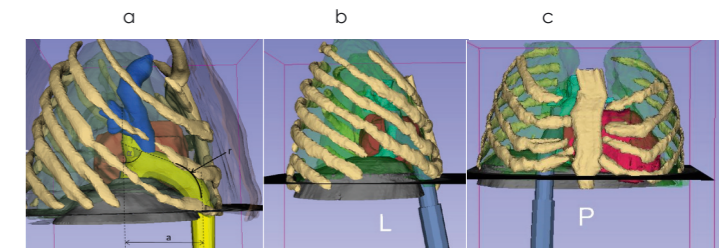


Abb. 2: Anpassbare Parameter der Kanüle (a), Ansicht der positionierten Kanüle (b und c)

Sieben Herzchirurgen werden hierzu in verschiedenen virtuellen Fittings zur Positionierung und Geometrie der Kanüle befragt und patientenspezifische Varianten ausgewählt. Dabei stellt sich heraus, dass für die Behandlung der diskutierten Patientenkohorte verschiedene Kanülengrößen notwendig sind.

Ebenfalls können die favorisierten Kanüengeometrien als 3D Druck-Teile als erste Muster hergestellt werden. Hierbei kommt der Formlabs 3D-Drucker mit dem Material Elastic 50A Resin zum Einsatz. Mit einer Shorehärte von 50A lässt sich ein Prototyp des Kanülenkopfes drucken, der den mechanischen Eigenschaften der Silikonkanüle ähnelt.

ERGEBNISSE UND DISKUSSION

Aufgrund der engen Zusammenarbeit aus Klinik (Bereitstellung der Bilddaten, Designabstimmung im virtuellen Fitting) und Industrie (Aufbereitung der Daten, Vorstellung verschiedener Kanüledesigns) konnte ein Produkt entwickelt werden, welches den besonderen anatomischen Gegebenheiten von Patienten mit Single-Ventrikel-Physiologie gerecht wird. Auch für zukünftige Patienten kann dieser Weg genutzt werden, um die passende Kanülengröße aus dem Produktportfolio auszuwählen.

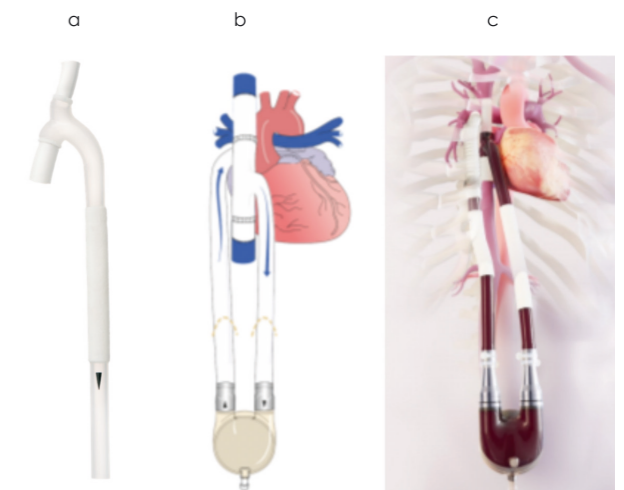


Abb. 3: Venöse Kanüle (a), Schemabild der implantierten Kanüle (b und c)

FAZIT

Durch die bis dato weltweit ersten fünf Implantationen der Kanüle zeigt sich, dass mit der venösen Kanüle in Kombination mit dem EXCOR Herzunterstützungssystem eine Überbrückung der Wartezeit bis zur Herztransplantation erreicht werden kann. Die Datenaufbereitung und das virtuelle Fitting führten zu einem direkten Austausch zwischen Anwendern und Entwicklern, wodurch Änderungswünsche und Ideen sehr schnell umgesetzt und bewertet werden konnten.

KI-gestützte Umsetzung für den digitalen Prozess von 3D-gedruckten Orthesen

Tino Jacobi (3D-Medico, Think3DDD), Amar Kumar Myadam (3D-Medico, Think3DDD), Youghourta Boutlikhet (3D-Medico), Noha Esselaoui (3D-Medico)

3D-Medico,
Rudower Chaussee 17, 12489 Berlin
eMail: info@3d-medico.de
<http://www.3d-medico.de>

Abstract. Dieser Beitrag beleuchtet die Herausforderungen, denen sich Patienten mit konventionellen Orthesen gegenübersehen, und präsentiert eine innovative, KI-gestützte Lösung, die mithilfe von 3D-Drucktechnologie personalisierte Orthesen produziert. Die Integration von KI in den Scanning- und Modellierungsprozess revolutioniert die Orthese-Herstellung und betont den Wert von ökologischer Nachhaltigkeit, Wirtschaftlichkeit und Patientenzufriedenheit.

1 EINLEITUNG UND HINTERGRUND

Die orthopädische Versorgung spielt eine entscheidende Rolle im Gesundheitssystem¹, wobei Bandagen und Orthesen eine zentrale Rolle bei der Behandlung und Rehabilitation von Verletzungen spielen. Dennoch stehen Patienten oft vor erheblichen Herausforderungen bei der Verwendung konventioneller Orthesenlösungen. Probleme wie Hautirritationen, Entzündungen, Wasserempfindlichkeit und eingeschränkte Atmungsaktivität der Materialien beeinträchtigen den Komfort und die Effizienz solcher traditionellen Hilfsmittel. Dies führt zu einem dringenden Bedarf an individualisierten, patientenspezifischen Lösungen, die sowohl effektiv als auch komfortabel sind.

In einer Zeit, in der Technologie und Medizin sich rasch weiterentwickeln, bietet die Integration von künstlicher Intelligenz (KI) und 3D-Druck eine vielversprechende Lösung für diese Probleme. 3D-Medico hat diesen Ansatz erkannt und sich zum Ziel gesetzt, den Prozess der Orthesen-Herstellung zu revolutionieren. Indem wir die Vorteile der KI sowohl im Scanning-Prozess als auch in der Modellierung nutzen, streben wir nicht nur nach einer verbesserten Patientenversorgung, sondern legen auch Wert auf ökologische Nachhaltigkeit und Wirtschaftlichkeit. In diesem Beitrag werden wir unsere innovative Methode vorstellen und diskutieren, wie sie den aktuellen Status quo in der orthopädischen Versorgung herausfordert und neu definiert.

2 KI IM 3D-SCAN-PROZESS

Der medizinische Bereich stellt besondere Herausforderungen an die Genauigkeit und Zuverlässigkeit von KI-Systemen. Fehler können direkte Auswirkungen auf die Gesundheit und das Wohlbefinden der Patienten haben. Daher ist es entscheidend, dass der 3D-Scan-Prozess mit höchster Präzision durchgeführt wird. Bei 3D-Medico nutzen wir eine bahnbrechende Methode, die die Kapazitäten moderner Smartphones maximiert. Ein medizinischer Fachmann zieht einen gemusterten Strumpf über den betroffenen Körperbereich des Patienten und platziert darauf Skalierungspunkte. Mithilfe eines Smartphones wird dann ein 360-Grad-Video aufgenommen und zur weiteren Verarbeitung an unsere spezialisierten Server übermittelt.

Unsere KI-Technologie tritt hier in Aktion. Sie verwendet fortschrittliche Algorithmen der Photogrammetrie, um aus diesem Video ein detailliertes 3D-Modell zu erstellen. Die KI bereinigt den Scan von jeglichen Störungen, indem sie unnötige Daten und Umgebungsobjekte eliminiert. Zusätzlich erkennt die KI die auf dem gemusterten Strumpf platzierten Skalierungspunkte. Dies sichert die genaue Proportionalität und Größe des Modells.

Die Vorteile für den Orthopädietechniker sind klar: Durch den Einsatz von KI im Scanning-Prozess sparen sie wertvolle Zeit und minimieren gleichzeitig die Fehleranfälligkeit. Zudem können sie sicher sein, dass die erhaltenen Modelle von hoher Qualität und Genauigkeit sind.

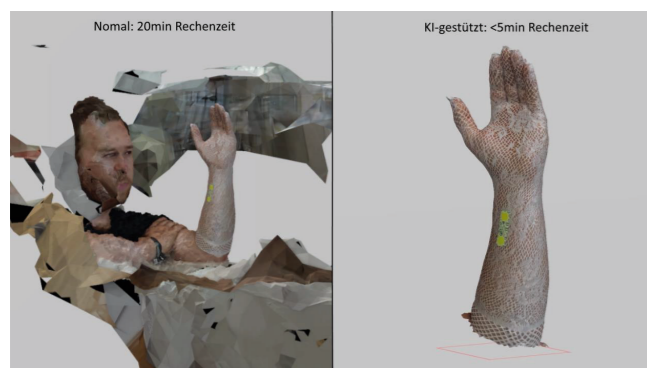


Abb. 1. Scanergebnis: normale Photogrammetrie (links) vs. optimierte Photogrammetrie mit KI (rechts)

3 KI-GESTÜTZTE MODELLIERUNG VON ORTHESEN

Der Prozess der Orthese-Modellierung ist komplex und erfordert sowohl medizinisches als auch technisches Verständnis. Insbesondere in der Medizintechnik ist es von entscheidender Bedeutung, dass die erstellten Modelle nicht nur funktional, sondern auch anatomisch korrekt sind. Die KI-Technologie von 3D-Medico revolutioniert diesen Prozess und macht ihn sowohl einfacher als auch präziser.

Zunächst identifiziert die KI spezifische anatomische Merkmale des gescannten Körperteils. Im Falle einer Hand, zum Beispiel, erkennt sie die Struktur der Hand, die Position und Länge der Finger und andere wichtige Merkmale. Auf Basis dieser Erkennung generiert die KI ein biomechanisches 3D-Skelett, das den Bewegungsapparat der Hand genau abbildet.

Das Besondere an diesem Skelett ist seine Anpassbarkeit. Anstatt von Grund auf ein neues Modell erstellen zu müssen, bietet die KI dem Orthopädietechniker ein vorgefertigtes, aber vollständig anpassbares Modell. Der Techniker kann nun wenn gewünscht gezielt Punkte des Skeletts verschieben. Danach kann er die Winkel und Rotationen der Gelenke individuell anpassen. Dies ist besonders nützlich, wenn Haltungskorrekturen erforderlich sind. Anstatt ein komplexes neues Skelett manuell zu erstellen, können Anpassungen mit wenigen Klicks vorgenommen werden. Dies spart nicht nur wertvolle Zeit, sondern reduziert auch das Risiko menschlicher Fehler.

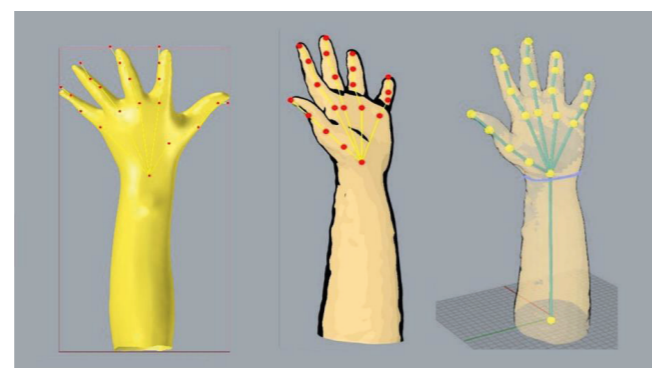


Abb. 2. KI-generiertes biomechanisches Skelett 2D (links) und 3D (rechts)

Während der Orthopädietechniker der Schritt-für-Schritt-Modellierung der Orthese folgt, optimiert die KI das Design auch für den 3D-Druckprozess. Sie analysiert das Orthesenmodell auf mögliche Problemstellen, wie Überhänge, und passt das Design entsprechend an. Dies stellt sicher, dass das Endprodukt sowohl medizinisch wirksam als auch druckbar ist.

Für den Orthopädietechniker bedeutet dies eine deutliche Vereinfachung des gesamten Designprozesses. Durch die Unterstützung der KI wird der Prozess nicht nur schneller, sondern auch intuitiver und weniger fehleranfällig. Der Techniker kann sich auf seine medizinische Expertise konzentrieren, während die KI die technischen Details übernimmt.

4 DER 3D-DRUCKPROZESS UND NACHHALTIGKEIT

Der 3D-Druck, insbesondere im medizinischen Bereich, hat sich in den letzten Jahren rasant entwickelt und bietet nun die Möglichkeit, individuell angepasste Produkte mit hoher Präzision herzustellen. Bei 3D-Medico verwenden wir Hochgeschwindigkeits-FFF-Drucker, die speziell für den Druck von medizinischen Produkten optimiert sind. Diese Drucker können flexible Materialien effizient und zuverlässig verarbeiten, was für die Produktion von Orthesen unerlässlich ist.

Ein besonderes Augenmerk wird auf die Auswahl der Materialien gelegt. Wir setzen auf umweltfreundliche Materialien wie PP, TPU. Diese Materialien sind nicht nur recycelbar, sondern auch für ihre Haltbarkeit und Funktionalität im medizinischen Bereich bekannt. Durch die Verwendung dieser Materialien und den anschließenden Recyclingprozess schaffen wir einen geschlossenen Produktionskreislauf, der die Umweltbelastung minimiert.

Unsere 3D-Drucker werden mit einem integrierten QM-System ausgestattet, das eine lückenlose und medizinkonforme Dokumentation des Produktionsprozesses gewährleistet. Dies stellt sicher, dass jede Orthese, die produziert wird, den festgelegten Qualitätsstandards entspricht und den Patienten optimal unterstützt.



Abb. 3. 3D-gedruckte halboffene Daumen-Hand-Orthese

5 FAZIT UND AUSBLICK

Die Integration von KI in den Prozess der Orthesenherstellung hat bereits erhebliche Verbesserungen in den Bereichen Scanning und Modellierung gebracht. Die mögliche zukünftige Anwendung von KI zur Überwachung des 3D-Druckprozesses könnte eine weitere Revolution darstellen. Eine solche KI würde nicht nur dazu beitragen, den Druckprozess zu optimieren, sondern auch die Qualität und Zuverlässigkeit jedes gedruckten Produkts sicherzustellen.

Ein weiterer spannender Ausblick ist die Integration der KI-Handerkennung in eine App, die den Nutzer bei der Videoaufnahme anleitet. Dies würde den Scanning-Prozess weiter vereinfachen und den Orthopädietechnikern die Möglichkeit geben, noch präzisere Modelle für den Druck zu erstellen.

Es ist jedoch wichtig zu betonen, dass einige der hier beschriebenen KI-Anwendungen noch in der Prototypenphase³ sind. Bei 3D-Medico sind wir bestrebt, diese Funktionen schnellstmöglich in unsere öffentlich verfügbare Software zu integrieren.



Abb. 4. Team 3D-Medico mit Orthesen: Amar, Noha, Youghourta, Tino (von links nach rechts)

In Bezug auf Herausforderungen wird es entscheidend sein, die regulatorischen Anforderungen⁴ und Standards für den Einsatz von KI in der Medizintechnik genau zu beobachten, insbesondere im Hinblick auf Datenintegrität und Zuverlässigkeit. Es ist jedoch unbestreitbar, dass KI das Potenzial hat, einige der am häufigsten verwendeten Standard-Orthesen zu automatisieren. Dies könnte die Orthopädietechnik, die bereits unter einem Fachkräftemangel leidet, erheblich entlasten und gleichzeitig die Qualität der Patientenversorgung erhöhen.

Bei 3D-Medico sind wir stolz darauf, an vorderster Front dieser Innovationen zu stehen und freuen uns darauf, weiterhin Lösungen zu entwickeln, die sowohl für medizinisches Fachpersonal als auch für Patienten von Vorteil sind.

LITERATURVERZEICHNIS

1. Ausgaben der Barmer GEK für Bandagen und Orthesen nach Anwendungsort im Jahr 2011, <https://de.statista.com/statistik/daten/studie/247379/umfrage/ausgaben-fuer-bandagen-und-orthesen-der-barmer-gek-nach-anwendungsorten/>, 30.10.2023
2. Video: Wie schnell kann ein 3D-Scan aufgenommen und verarbeitet werden?, <https://3d-medico.de/3d-scan.html>, 30.10.2023
3. Video: 3D-Medico Beta-Software zur Generierung der Orthimale Vorschau, <https://3d-medico.de/3d-software.html>, 30.10.2023
4. Regulatorische Anforderungen an Medizinprodukte mit Machine Learning, <https://www.johner-institut.de/blog/regulatory-affairs/regulatorische-anforderungen-an-medizinprodukte-mit-machine-learning/>, 30.10.2023

Evaluation of Depth Completion Using Solid-State LiDAR Dataset

Ying Yin Li, Jan Herschel, Nick Theisen,
Roman Abayev, Dietrich Paulus, Anselm von Gladiss

Arbeitsgruppe Aktives Sehen, Institut für Computervisualistik,
Universität Koblenz, Universitätsstraße 1, 56070 Koblenz
liyingyinsimon@uni-koblenz.de, jherschel@uni-koblenz.de,
nicktheisen@uni-koblenz.de, abayev@uni-koblenz.de,
paulus@uni-koblenz.de, vongladiss@uni-koblenz.de

Abstract. This paper aims to explore depth completion using sparse depth from a solid-state LiDAR sensor. With many practical applications such as autonomous driving, 3D reconstruction, medical imaging, accurate dense depth is needed. State-of-the-art neural networks designed for depth completion on Velodyne datasets are utilized and adapted for the solid-state LiDAR dataset collected at Universität Koblenz. We use a dataset with ground truth for training the neural network model and validating the results. We use sparse depth as the only input as RGB images may not be available in general. Ground truth data is generated by capturing multiple shots of the same scene. The objective is to assess the potential of existing algorithms when applied to solid-state LiDAR sensors. The dataset is evaluated using various methods, and the results are compared and analyzed.

1 BACKGROUND AND MOTIVATION

Computer vision and image processing have been long fascinating and challenging fields of research. Many applications nowadays including autonomous driving, medical imaging, agriculture, robotics etc. have greatly improved human lives. Depth completion is one of the computer vision tasks. It is the process of estimating the dense depth information of a scene from a sparse depth point cloud data.

KITTI dataset [Uhr+17] and similar datasets have been widely used in computer vision research, especially in the field of autonomous driving, for evaluating and developing algorithms using Velodyne LiDAR data. Solid-state LiDAR sensors, on the other hand, are a more recent advancement in LiDAR technology. Unlike traditional mechanical spinning LiDAR sensors like those produced by Velodyne, solid-state LiDAR sensors do not have moving parts. In this paper, we are going to explore the potential of using solid-state LiDAR sensors for depth completion.

RGB images are often used as a guidance in depth completion as it provides more geometric information like edges and fine details that sparse depth does not have. However, RGB images may not be available in general or there may be some limitations i.e. very dark environment or hardware limitations.

Sparse 3D point cloud is projected on a 2D space. Different scenes of 2D sparse depth contain missing depth values which needed to be filled. Depth completion is the task that of predicting a dense depth image from

sparse depth data. In this study, only sparse depth data is used as input. To quantitatively evaluate the results of depth completion, ground truth data is needed. Manual annotation of ground truth is hard to acquire. For our solid-state LiDAR [AGP21] dataset, ground truth data is obtained by integrating multiple scans from the same scene assuming the scenes are static.

In recent years, deep learning-based methods for various computer vision tasks become popular. It uses convolution neural networks (CNNs) [Gu+18] to learn discriminative features such as edges, shapes, textures for accurate and efficient segmentation. Using CNNs on RGB images for various computer vision tasks has been well developed. However, in depth completion task, sparse depth measurement is used as input which lack of clear features and geometric information compared to RGB images. Different neural network architectures and modules are developed for depth completion. In this study, some of the neural networks are chosen for the depth completion tasks using solid-state LiDAR dataset. The results are compared and analyzed.

2 METHODS

The solid-state LiDAR dataset utilized in this study comprises 145 distinct scenes and is obtained from the Livox Mid-100 (CHN) dataset encompassing 114 training images and 31 validation images. The input consists of multiple sparse depth maps consecutively captured by a Livox Mid-100 LiDAR, utilizing a narrow integration window (i.e., 0.03374 seconds) for each frame. The experiment involves collecting 3D point cloud data, which is then transformed into a 2D image space. The image size is 1250x375. Due to the non-repetitive trajectories of the Livox Mid-100 sensor, it becomes feasible to generate ground truth data by integrating multiple scans from the same scene. Over 50 frames are utilized for the ground truth data. In addition, the background or far away object are not under consideration i.e. the maximum depth is constrained to 100 meters.

The „KITTI Vision Benchmark Suite“ [Gei+13] has been widely used for various computer vision tasks including semantic segmentation, depth completion, object detection etc. It consists of over 93,000 depth maps from Velodyne HDL-64E and corresponding RGB images. It has been widely used for evaluation of depth completion. The neural networks chosen in this paper also tes-

ted on KITTI dataset. We adapt these neural networks to our solid-state LiDAR dataset to see if they can achieve a similar result.

The Virtual KITTI Depth Dataset [Gai+16] is a synthetic dataset generated using the Unity game engine. It aims to simulate the conditions of the KITTI Depth dataset and contains over 21,000 images of outdoor scenes. The dataset includes corresponding depth maps, ground truth labels, and camera poses. It is used in the training of ScaffNet [WCS21] since it has more data for training and it has dense depth maps as ground truth. We can simulate the trajectories of solid-state LiDAR for generating sparse depth input. For the depth completion task, we employ the neural network ScaffNet [WCS21], which represents a state-of-the-art approach capable of depth maps completion using sparse 2D depth images exclusively. In order to provide a comprehensive comparison, we also evaluate the performance of Deep Convolutional Compressed Sensing (DCCS) [CWL19] on the solid-state LiDAR dataset.

DCCS is a framework that synergizes compressed sensing and deep learning for depth completion exclusively utilizing sparse depth measurements. Compressed sensing is a signal processing technique designed for reconstructing sparse signals. Given that an image is essentially a signal, it is feasible to reconstruct images from sparse image data. This technique can also be extended for depth completion of sparsely sampled depth maps. Moreover, the network proposed in the paper is built upon the foundation of the Alternating Direction Neural Network (ADNN) [MCL18].

ScaffNet is a neural network that uses Spatial Pyramid Pooling (SPP) [He+15] and encode-decoder network. SPP is a module used between convolution layers and fully connected layers. The output feature map of convolution layers can have arbitrary sizes. After the spatial pyramid pooling layer, the output features are concatenated to a fixed-length representation. The SPP module is modified and added to densify the sparse inputs from sparse point clouds. When using pooling with small kernels, details are preserved, but very few non-zero activations are produced in the subsequent layers. When using large kernels, the sparse inputs are densified, but the details are lost. Concatenating the results from small and large kernels with suitable weightings produces a good balance between density and details. In the

encoder-decoder part, the encoder utilizes five convolution layers which are similar to the VGG block [SZ14]. The outputs of each convolution layer are fed into a decoder as skip connections. The decoder consists of five modules. The outputs from the previous decoder module are deconvoluted to do the upsampling. Then the output is concatenated with the output of the corresponding resolution (skip connection) from the encoder. The output undergoes a convolution with a 3x3 kernel, and the result is passed into the next decoder module.

Firstly, ScaffNet is trained using our training dataset. Secondly, it is also trained using the Virtual KITTI dataset, similar to the approach described in [WCS21]. However, in our case, the sparse input data relies on the trajectories captured by the Livox Mid-100 LiDAR sensor instead of the Velodyne HDL-64E LiDAR sensor. Due to the different input patterns of the Livox Mid-100 LiDAR sensor, adjustments need to be made to the training data to achieve more accurate results. In the case of the ground truth for the Virtual KITTI dataset, a dense depth ground truth is used, following the methodology outlined in [WCS21]. In the case of DCCS, the neural network is trained using our training dataset.

3 RESULTS

The experiment were conducted in order to see how the architectures of different neural networks behave on the solid-state LiDAR dataset. In [WCS21], ScaffNet model was trained with Virtual KITTI dataset and it gives a promising result in the evaluation of KITTI dataset. A similar approach is used in our experiment. However, the sparse depth input trajectories are changed to match the solid-state LiDAR sensor trajectories. For DCCS, the model is trained with our solid-state LiDAR training dataset.

Fig. 1 displays the sparse depth input image, ground truth, and predictions for various methods for different binnings. The experimental results are summarized in Table 1. The different binnings in time represent various levels of sparsity in the input sparse data. In order to evaluate the performance of the depth completion algorithms, two widely used error metrics, mean absolute error (MAE) and root mean square error (RMSE), are employed. The metrics are reported in millimeters.

The MAE is calculated as the average absolute difference between the ground truth and predicted values,

and it is computed using the following formula:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|$$

where y_i represents the ground truth pixel value at the i -th pixel position, x_i represents the pixel value obtained from the result of different algorithms at the i -th pixel position, and n is the total number of valid pixels. The RMSE is calculated as the square root of the average squared difference between the ground truth and predicted values, and it is computed using the following formula:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}$$

Based on the results, it is evident that ScaffNet outperforms DCCS in terms of accuracy for depth completion. Furthermore, within the ScaffNet algorithm, the model trained using the Virtual KITTI dataset demonstrates superior results compared to the one trained using the Livox dataset. Additionally, the findings indicate that increasing the number of frames in the sparse input generally leads to more accurate prediction results. Regarding the DCCS algorithm, the results indicate that denser input data leads to better metric performance.

4 DISCUSSION

From the results, it can be observed that ScaffNet with a pretrained model on the Virtual KITTI dataset performs the best among the algorithms tested.

For training a neural network, we provide a dataset with ground truth. However, since the dataset we collected may be smaller in size compared to publicly available datasets, the resulting performance may not be fully representative. In such cases, it is beneficial to use the dataset with synthetic data, such as the Virtual KITTI dataset, which provides a large number of samples and dense ground truth depth maps. By subsampling the Virtual KITTI dataset, we can simulate various solid-state LiDAR trajectories and enhance the training process.

From the results obtained using different binnings, it is evident that higher binnings yield more accurate results. This can be attributed to the reduced sparsity, as a higher level of spatial and geometric information is available in the input data. Consequently, depth completion becomes easier due to the increased amount of information.

However, it is important to note that there is a trade-off involved. With higher binnings of frames, more time is required to capture the data, which can impact real-time processing capabilities. Therefore, achieving a balance between accuracy and real-time performance is crucial when considering the choice of binnings for depth completion tasks.

Additionally, it is observed that for both methods, there is a substantial improvement from binnings of 1 to 5. However, the improvement becomes less significant with higher binnings. Considering this, binnings of 5 would be a preferable choice for real-time processing, as it strikes a balance between accuracy and processing efficiency.

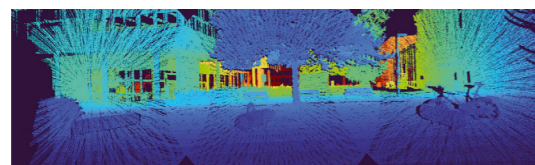
5 CONCLUSION

To conclude, we apply different neural networks on solid-state LiDAR dataset for depth completion and compare the results. The introduction of solid-state LiDAR technology brings new opportunities and challenges for depth completion and scene understanding. The availability of more datasets in the future, specifically incorporating solid-state LiDAR data, coupled with the development of advanced algorithms for depth completion, holds great promise for advancing the field of computer vision.

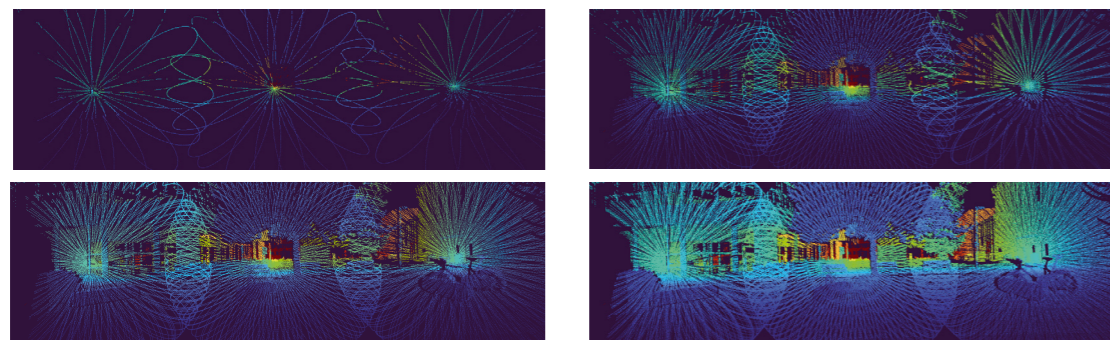
Method	Livox			
	MAE (mm)			
Binnings time in seconds	0.03374	0.1687	0.3374	0.8435
Number of frames to be integrated over time	1	5	10	25
ScaffNet [WCS21] pretrained with Virtual KITTI dataset with Livox pattern	1368.745	734.572	611.567	543.317
ScaffNet [WCS21] pretrained with Livox dataset	1726.711	1190.231	1050.913	1053.380
DCCS [CWL19]	6563.293	2149.185	1898.639	1630.135

Method	RMSE (mm)			
	RMSE (mm)			
Binnings time in seconds	0.03374	0.1687	0.3374	0.8435
Number of frames to be integrated over time	1	5	10	25
ScaffNet [WCS21] pretrained with Virtual KITTI dataset with Livox pattern	4509.079	3039.113	2676.120	2401.45
ScaffNet [WCS21] pretrained with Livox dataset	4572.049	3820.317	3544.825	3522.841
DCCS [CWL19]	11441.495	4651.383	4224.370	3738.646

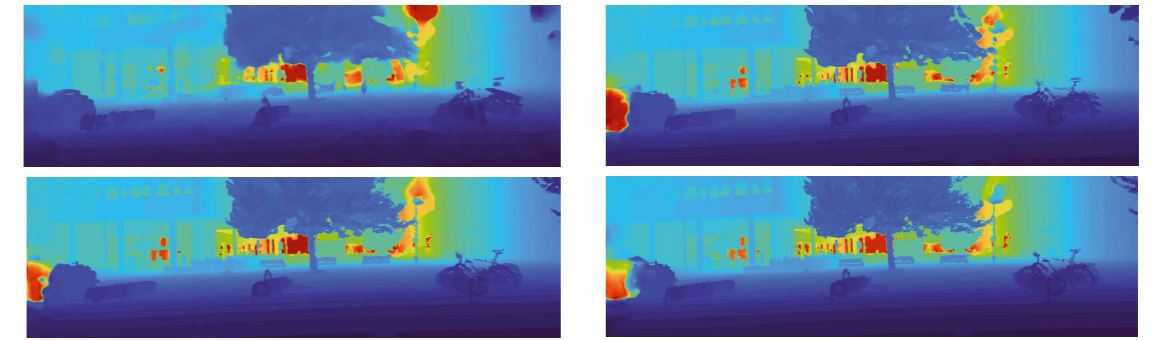
Table 1. Evaluation results on different algorithms on Livox dataset



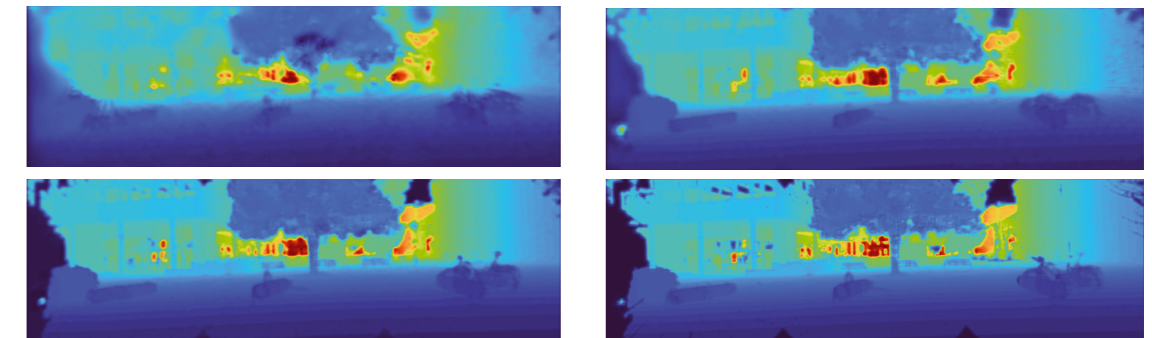
(a) Ground truth



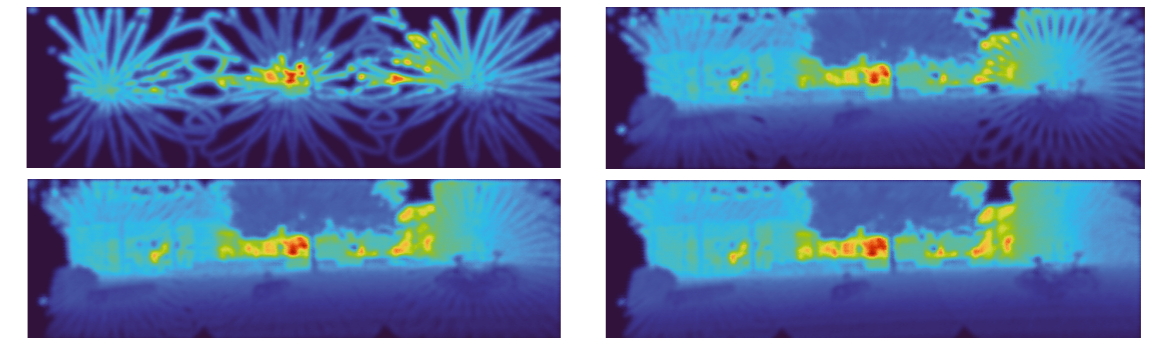
(b) Sparse depth input



(c) ScaffNet [WCS21] pretrained with Virtual KITTI dataset



(d) ScaffNet [WCS21] pretrained with Livox dataset



(e) DCCS [CWL19]

Fig. 1. Sample result showing depth completion of Livox using the „turbo“ colormap. From left to right, and top to bottom, the subfigures correspond to a binning of 1, 5, 10, and 25.

REFERENCES

- [Gei+13] Andreas Geiger et al. "Vision meets Robotics: The KITTI Dataset". In: International Journal of Robotics Research (IJRR) (2013).
- [SZ14] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: arXiv preprint arXiv:1409.1556 (2014).
- [He+15] Kaiming He et al. "Spatial pyramid pooling in deep convolutional networks for visual recognition". In: IEEE transactions on pattern analysis and machine intelligence 37.9 (2015), pp. 1904–1916.
- [Gai+16] Adrien Gaidon et al. "Virtual worlds as proxy for multi-object tracking analysis". In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2016, pp. 4340–4349.
- [Uhr+17] Jonas Uhrig et al. "Sparsity Invariant CNNs". In: International Conference on 3D Vision (3DV). 2017.
- [Gu+18] Jiuxiang Gu et al. "Recent advances in convolutional neural networks". In: Pattern recognition 77 (2018), pp. 354–377.
- [MCL18] Calvin Murdock, MingFang Chang, and Simon Lucey. "Deep component analysis via alternating direction neural networks". In: Proceedings of the European Conference on Computer Vision (ECCV). 2018, pp. 820–836.
- [CWL19] Nathaniel Chodosh, Chaoyang Wang, and Simon Lucey. "Deep convolutional compressed sensing for lidar depth completion". In: Computer Vision—ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part I 14. Springer. 2019, pp. 499–513.
- [AGP21] Roman Abayev, Anselm von Gladiss, and Dietrich Paulus. "Eigenschaften verschiedener Solid-State-LIDAR-Systeme". In: 3D-NordOst 2021. Ed. by St. Rothstock et al. Berlin, Germany: Gesellschaft zur Förderung angewandter Informatik e.V. (GFaI), 2021, pp. 19–27. ISBN: 978-3-942709-27-9. URL: <https://www.uni-koblenz.de/~agag/Public/Abayev2021EVS.pdf>.
- [WCS21] Alex Wong, Safa Cicek, and Stefano Soatto. "Learning topology from synthetic data for unsupervised depth completion". In: IEEE Robotics and Automation Letters 6.2 (2021), pp. 1495–1502.

Entwicklung eines Verfahrens zur auflösungsangepassten Abschätzung der Schichtdicke in einer modellbasierten Auftragssimulation durch adaptive Granularisierung

Leonardo Maben, Stephan Brodkrob, Benjamin Hohnhäuser

Gesellschaft zur Förderung angewandter Informatik e.V.
Bildverarbeitung / Industrielle Anwendungen
Volmerstraße 3, 12489 Berlin
www.gfai.de

Abstract. Dieses Projekt konzentriert sich auf die Entwicklung einer Methode zur präzisen Abschätzung der Schichtdicke in Beschichtungsprozessen. Dabei werden verschiedene Remeshing-Methoden angewandt und miteinander verglichen, um die Genauigkeit der Schichtdickenbestimmung zu verbessern. Das Ziel besteht darin, herauszufinden, welche Remeshing-Methode die genauesten Ergebnisse für die Schichtdickenabschätzung liefert. Dieses Projekt zielt darauf ab, die Qualität und Effizienz von Beschichtungsprozessen in der industriellen Fertigung zu steigern, indem die Genauigkeit von Vorhersagen und Optimierungen erhöht wird. Mithilfe dieser Methoden werden zwei Sprühverfahren für eine Auftragssimulation entwickelt, um genauere Ergebnisse zu erzielen.

1 EINLEITUNG

In der industriellen Serienfertigung haben automatisierte Beschichtungsprozesse in den letzten Jahrzehnten kontinuierlich zugenommen. Die Beschichtung dient dabei sowohl als dekorativer Zusatz als auch funktionalen Zwecken, wie z.B. dem Schutz vor Korrosion. Allgemein kann davon ausgegangen werden, dass der Anteil der manuellen Beschichtungsprozesse in kleinen Betrieben höher ist als bei der Serienfertigung in Großunternehmen. Das Fehlen von automatisierten Roboteranwendungen in diesem Bereich kann darauf zurückgeführt werden, dass es wirtschaftlich nicht darstellbar ist, manuelle Roboterprogramme für kleine Losgrößen zu erstellen. Ein weiterer Grund ist die Tatsache, dass die Beschichtungsergebnisse nicht im Voraus sicher vorhergesagt oder bewertet werden können. Manuelle Beschichtung bedeutet in diesem Fall eine nicht gleichbleibende Qualität und eine größere Anzahl nicht fachgerecht beschichteter Werkstücke im Verhältnis zur Lackierung durch Roboter. Fehlerhaft beschichtete Werkstücke führen dazu, dass bei einer Abweichung von den gewünschten Ergebnissen diese im besten Fall nachbearbeitet bzw. neu beschichtet werden müssen. Im schlimmsten Fall sind sie Ausschuss. Hierdurch entstehen hohe Kosten, Materialverlust und eine verlängerte Arbeitsdauer durch die Neubeschichtung des Werkstücks.

Eine Möglichkeit, den wirtschaftlich sinnvollen Einsatz von Robotern zu ermöglichen, ist die Verwendung einer Beschichtungssimulation zur Vorhersage der Beschichtungsergebnisse. Dadurch kann der Beschichtungsprozess in der Simulation optimiert werden. Aufgrund der hohen Komplexität bei den Berechnungen für die Simulation des gesamten Beschichtungsprozesses ist es notwendig, durch die Verwendung von Näherungswerten

und Abschätzungen, den Berechnungsaufwand in handhabbaren Größen zu halten. Diese Näherungen führen zu einer ungenauen Berechnung der Beschichtungsergebnisse und der Schichtdicke. Bei Kleinserien bzw. kleinen Losgrößen und insbesondere bei Losgröße 1 würde sich der Einsatz einer besseren Beschichtungssimulation, die zu einer besseren Abschätzung der Ergebnisse führt und die auftretenden Modellfehler reduziert, als enorm nützlich erweisen.

Das Ziel des Projekts ist es, ein neues Verfahren für eine bestehende Lösung zur modellbasierten Schichtdicken-simulation zu entwickeln, welches die Abschätzung des tatsächlichen Modellfehlers verbessert. Dabei soll die Verwendung von Heuristiken zur Berechnung von sicheren Ober- und Untergrenzen bei der Beschichtungssimulation für jedes Dreieck des Oberflächenmodells und des gesamten Werkstücks adressiert werden.

2 WAS IST REMESHING?

Ein 3D-Mesh oder 3D-Netz ist der strukturelle Aufbau eines 3D-Modells, das aus Polygonen besteht. 3D-Netzen bestehen aus Referenzpunkten, deren Koordinaten in X, Y und Z angegeben werden, um Formen mit Höhe, Breite und Tiefe zu definieren. Das 3D-Netz kann manipuliert werden, um Bilder, Animationen und Simulationen zu erstellen. 3D-Netzen haben mehrere Vorteile gegenüber anderen 3D-Modellierungstechniken. Sie sind relativ einfach zu beschreiben, skalierbar und effizient, wodurch verschiedene Formen von einfachen Würfeln bis hin zu komplexen Figuren erstellt werden können.

Häufig sind diese Netze aufgrund von Unebenheiten, Unstetigkeiten und inhomogener Verteilung der Stützpunkte für die weitere Bearbeitung ungeeignet und es besteht die Notwendigkeit, die Qualität zu verbessern. Remeshing ist der Prozess, bei dem aus einem 3D-Netz eine bessere diskrete Darstellung der zugrundeliegenden Oberfläche gefunden wird. [1] Bei der Erörterung von Remeshing-Techniken wird zumeist die lokale Struktur des Objekts betrachtet. Die lokale Struktur eines Meshes wird durch **den Typ, die Form und die Verteilung** der Mesh-Elemente beschrieben.

Der Typ bezieht sich auf die Art des Polygons, das für die Konstruktion des Meshes verwendet wird. Die am häufigsten verwendeten Zieltypen sind Dreiecke und Vierecke. Dreieck-Meshes sind in der Regel einfacher zu erstellen.

Das liegt daran, dass Computergrafiken normalerweise mit den Eckpunkten arbeiten. Für jedes Dreieck muss das System mit drei Eckpunkten pro Dreieck arbeiten, im Gegensatz zu einem Quadrat, das mit vier Eckpunkten pro Quad arbeitet. Ein großes Mesh kann aus mehreren Polygonen bestehen, die sich an einem einzigen Eckpunkt kreuzen. Mehrere Dreiecke können einen gemeinsamen Eckpunkt haben, während bei einem Quad-Mesh nur vier Quadranten einen gemeinsamen Eckpunkt haben. Dadurch verringert sich natürlich die Rechenzeit, die für die Erzeugung von mehr Dreiecken in einem Mesh im Vergleich zu Quadranten erforderlich ist. [1] Für die Durchführung dieses Projekts werden Dreiecks-Meshes verwendet. Abb.1 stellt grafisch die Unterschiede der beiden Elementtypen dar, wobei die erste Struktur aus Dreiecken und die zweite aus Vierecken besteht.

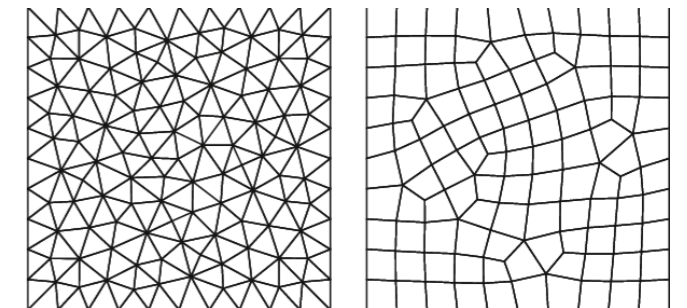


Abb. 1 Dreieck-Mesh und Quad-Mesh [2]

Die Form der Elemente kann in zwei Typen unterteilt werden: isotrop und anisotrop. Isotrope Dreiecke sind in alle Richtungen lokal einheitlich. Im Idealfall ist ein Dreieck isotrop, wenn es annähernd gleichseitig ist.

Anisotrope Dreiecke hingegen variieren je nach Ausrichtung auf der Oberfläche. Dies führt zu einer besseren Formnäherung mit einer geringeren Anzahl von Dreiecken. Der Kompromiss besteht in Meshes, die aus ungleichmäßigen Dreiecken bestehen, um die geometrische Struktur auszudrücken [3]. Die Abb. 2 stellt die beiden Formen der Elemente dar. Die Elemente auf der linken Seite in Abb. 2 sind stark anisotropisch, während die Elemente auf der rechten Seite eine hohe Isotropie aufweisen.

Die Verteilung der Elemente kann als einheitlich oder uneinheitlich klassifiziert werden. Dabei bezieht sich, ersteres auf ein Mesh, das Dreiecke enthält, die gleichmäßig über das gesamte Modell verteilt sind. Letzteres enthält ein Mesh, bei dem die Anzahl der Elemente variiert, z. B. werden Bereichen mit starker Krümmung kleinere Elemente zugewiesen und erhöhe damit lokal die Dichte.

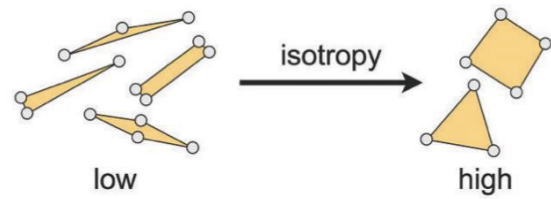


Abb. 2 Anisotrope vs. isotrope Elemente [1]

Durch die oben genannten Operationen wird die Dreiecksqualität erhöht und die Kantenoperationen dienen dazu, die Regelmäßigkeit des Meshes zu optimieren. Bei dieser Operation ist das Verschieben von Eckpunkten am aufwendigsten, da mittels Parametrisierung eine neue Eckpunktposition auf dem Mesh gefunden werden muss [4].

3 DAS REMESHING VERFAHREN

Für die Durchführung dieses Projekts werden drei besondere Methoden durch die Verwendung lokaler Remeshing-Operatoren, wie in Abb. 3 dargestellt, in Betracht gezogen. Die Remeshing-Operatoren können in vier Varianten unterteilt werden.

Edge Collapse wird in Situationen verwendet, in denen die Eckpunkte zweier Dreiecke so nahe beieinander liegen, dass sie in ihrem Mittelpunkt zu einem Eckpunkt zusammengefasst werden können. Auf diese Weise kann man die Anzahl der Dreiecke in dem gesamten Mesh reduzieren.

Edge Split ist das genaue Gegenteil, wobei zwei Dreiecke durch vier neue Dreiecke ersetzt werden, indem ein Eckpunkt hinzugefügt wird. Standardmäßig wird dieser Eckpunkt auf der ursprünglichen Kante platziert, so dass ein Split der einzige Operator ist, der die Maschenform garantiert nicht verändert.

Ein **Edge Flip**, manchmal auch als „Kantenrotation“ genannt, ersetzt zwei Dreiecke durch zwei neue Dreiecke. Zur Verbesserung der Dreiecksqualität werden Kanten gekippt, wenn dadurch das Minimum der sechs an dieser Operation beteiligten Winkel erhöht werden kann.

Ein **Vertex Shift** hingegen fügt keine neuen Dreiecke hinzu oder ersetzt sie, sondern reorganisiert Eckpunkte, sodass die Eckpunkte aus besseren Eigenschaften relativ

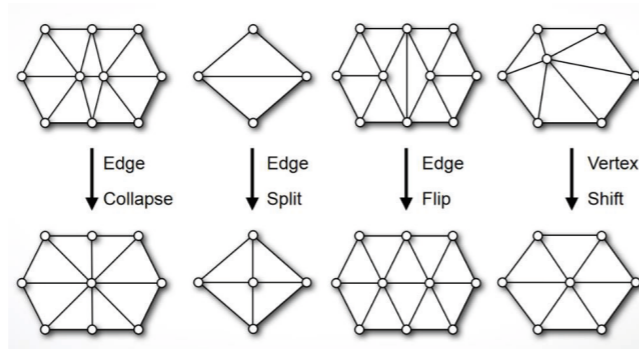


Abb. 3 Lokale Remeshing Operatoren

zu ihren Nachbarn bestehen. Die Eigenschaften entsprechen dabei Faktoren, wie beispielsweise ob die Dreiecke gleichseitig sind. [4]

Mit Hilfe dieser Remeshing-Operatoren werden die drei Remeshing-Methoden definiert, die zur Bildung einheitlicher Dreiecksnetze verwendet werden. Um diese Verfahren zu testen, wird zunächst ein 3D-Modell, wie in Abb. 4 zu sehen, als Eingabedaten verwendet, die später bearbeitet werden.

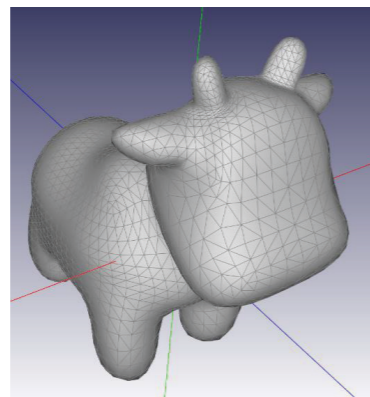


Abb.4 Ursprüngliches Testmodell

Direktes Remeshing: Die erste Methode beinhaltet ein direktes Remeshing-Verfahren, das Edge Collapse und Edge Split beinhaltet, um sicherzustellen, dass die Kantenlänge eines bestimmten Dreiecks immer gleich oder unter einem bestimmten Grenzwert ist. Dabei wird die Seite eines Dreiecks so lange geteilt, bis alle Dreiecke innerhalb der vorgegebenen Grenze liegen. Dies führt zur

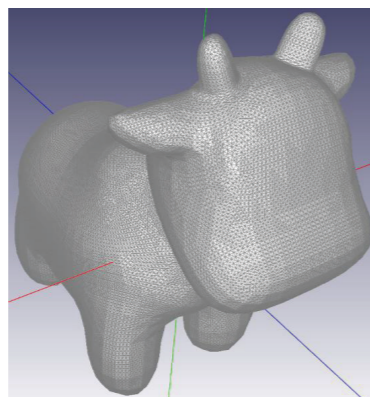


Abb. 4 Direktes Remeshing

Bildung eines gleichmäßigen Netzes. Die Abb. 5 zeigt das resultierende Dreiecksnetz, nachdem das Objekt dieser Art der Remeshing unterzogen wurde.

Gleichmäßiges Remeshing mit Delaunay-Triangulation:

Bei dieser Methode werden alle 4 Remeshing-Operatoren zusammen mit dem geometrischen Konzept der Delaunay-Triangulation verwendet, um Dreiecksnetze zu erzeugen. Unter Triangulation versteht man das Verbinden von Punkten zu Dreiecken, um daraus ein Netz zu bilden. Die Delaunay-Triangulation ist eine spezielle Art der Triangulation, die durch ihren Algorithmus regelmäßiger Dreiecke ergibt. Sie kann als eine Menge von Dreiecken definiert werden, die aus einer diskreten Menge von Eckpunkten besteht, so dass kein Eckpunkt innerhalb des Umkreises eines beliebigen Dreiecks der Menge liegt. Die Abb. 6 zeigt grafisch, wie dieses Triangulation mittels zwei Dreiecke aussieht.

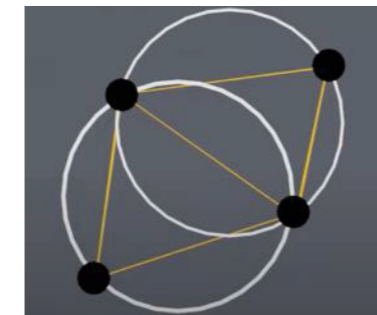


Abb. 6 Die Delaunay Triangulation mit den gegebenen Punkten

Der Algorithmus unterteilt eine Fläche optimal, indem er Punkte miteinander verbindet, um Dreiecke zu bilden, die die Eigenschaft des leeren Kreises erfüllen. Dadurch bleibt die Form erhalten und die Qualität des Dreiecksnetzes verbessert sich.

Die Abb. 7 zeigt das resultierende 3D-Netz nach Anwendung dieses Verfahrens.

Adaptive Remeshing: Adaptive Remeshing passt die Dichte des Dreiecksnetzes auf der Grundlage lokaler Merkmale an, verfeinert komplexe Regionen und vergrößert einfachere, wodurch Effizienz und Genauigkeit bei Simulationen und Rendering verbessert werden. Das Besondere an dem adaptiven Verfahren ist die Verwendung einer Ober- und Untergrenze für die Kantenlänge. Dies bedeutet, dass die Möglichkeit besteht, ein bestimmtes Netz mit gleichförmiger Dreiecke

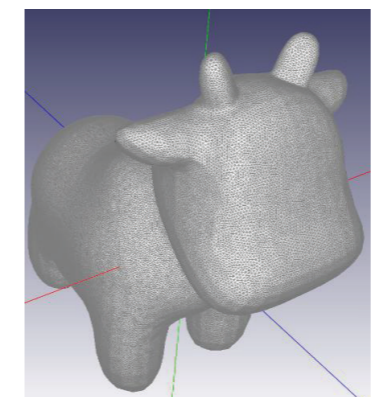


Abb.7 Remeshing mit Delaunay Triangulation

schneller zu erstellen, allerdings besteht auch das Risiko, dass sich die Gesamtform des Objekts ändert, wie in Abb. 8 dargestellt ist.

Zunächst wurden die drei Remeshing-Verfahren mit unterschiedlicher Kantenlänge und Anzahl der Dreiecke gegeneinander getestet. Mit der Verringerung der maximalen Kantenlänge eines Dreiecks nimmt auch die Anzahl der Dreiecke in dem neu gebildeten Netz zu. Die folgende Tabelle zeigt den Vergleich der einzelnen Verfahren.

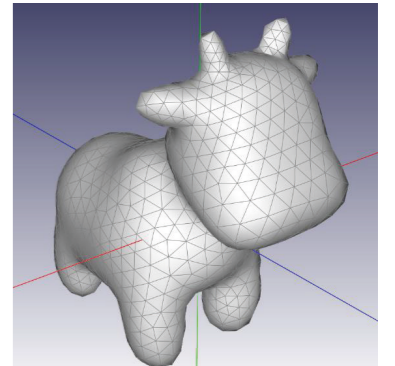


Abb. 8 Adaptive Remeshing

Es fällt sofort auf, dass das direkte Verfahren zwar ein gleichmäßiges Netz liefert, aber mit zunehmender Anzahl der Dreiecke der Zeitaufwand für das Remeshing massiv ansteigt. Im Vergleich dazu bieten die Delaunay- und die adaptive Methode eine viel schnellere Remeshing-Lösung für die Erzeugung einer großen Anzahl von Dreiecken. Obwohl die adaptive Lösung schnellere Ergebnisse liefert, besteht die Möglichkeit, dass die Netze ihre ursprüngliche Form nicht beibehalten.

Edge Length	Anzahl der Dreiecke	Direkt	Delaunay	Adaptive
0.05	14876	0,36	0,33	0,23
0.03	39904	1,48	0,77	0,77
0.02	85990	5,25	1,35	1,3
0.0015	157702	16,25	3,45	3,4
0.01	341586	79,47	4,08	3,94

4 ARTEN DER SPRÜHSIMULATION

Schwerpunkt-basierte Simulation: Die Annäherung an die Frage, ob ein bestimmter Bereich eines Dreiecksnetzes beschichtet wird oder nicht, ergibt sich aus der Positionierung der Schwerpunkte der Dreiecke, aus denen das Netz besteht. Wenn sich der Schwerpunkt des Dreiecks innerhalb des Sprühkegels befindet, wird das gesamte Dreieck beschichtet.

Eckpunkt-basierte Simulation: Die Annäherung an die Frage, ob ein bestimmter Bereich eines Dreiecksnetzes, genauer gesagt eines einzelnen Dreiecks, beschichtet wird, hängt von der Platzierung der Eckpunkte im simulierten Sprühkegel ab. Bei dieser Methode ist die Anzahl der Eckpunkte, die sich innerhalb des Sprühkegels befinden, der Parameter, der bestimmt, ob das Dreieck beschichtet wird oder nicht.

Um die Unterschiede zwischen den beiden Methoden unter Berücksichtigung ihrer jeweiligen Parameter zu evaluieren, wird ein Querschnitt aus dem zuvor erwähnten 3D-Modell genutzt. In diesem Ansatz werden bestimmte Dreiecke gezielt mit Sprühfarbe behandelt, wobei andere Einflussfaktoren, wie der Abstand zwischen der Sprühpistole und dem zu besprühenden Objekt sowie die Anzahl der Dreiecke, konstant gehalten werden. Das gewählte Muster wird gleichmäßig über die linke und rechte Seite des betrachteten Punktes verteilt, wodurch eine systematische Beobachtung des gesprühten Dreiecks ermöglicht wird.

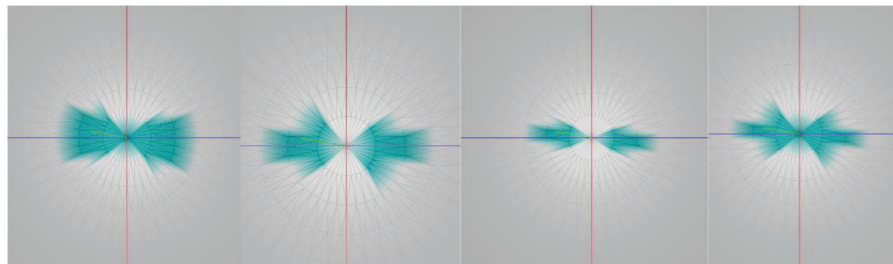


Abb. 9 Von links: Eckpunkt-basiertes Sprühen mit einem Eckpunkt, mit zwei Eckpunkten, mit drei Eckpunkten und Schwerpunkt-basiertes Sprühen

Die Unterschiede zwischen den Ergebnissen der folgenden Verfahren lassen sich anhand der Abb. 9 veranschaulichen. Der Querschnitt, beginnend mit dem linken Bild, ist eine eckpunkt-basierte Sprühsimulation, bei der nur ein Eckpunkt berücksichtigt wird. Wie auf dem Bild zu sehen ist, ist die Anzahl der gesprühten Dreiecke bei diesem Verfahren viel größer als bei den anderen. Im zweiten Bild wird eine eckpunkt-basierte Simulation mit zwei Eckpunkten präsentiert, die sich marginal von dem vorherigen Fall unterscheidet. Anschließend wird ein weiteres Bild gezeigt, das eine eckpunkt-basierte Simulation mit drei Eckpunkten darstellt und einen signifikanten Unterschied aufweist, da hierbei eine reduzierte Anzahl von zu versprühenden Dreiecken zu verzeichnen ist. Das abschließende Bild veranschaulicht eine schwerpunkt-basierte Sprühsimulation, die als eine Art Durchschnitt der zuvor präsentierten drei Fälle erscheint.

Die Analyse der Ergebnisse zeigt, dass die Sprühmethode mit einem Eckpunkt die Anzahl der zu besprühenden Dreiecke überschätzt, was zu einer größeren Anzahl von besprühenden Dreiecken führt, während die Sprühmethode mit drei Eckpunkten im Gegensatz zur vorherigen Methode die Anzahl der zu besprühenden Dreiecke unterschätzt. Dies führt zu einer Ober- und Untergrenze, wobei die tatsächliche Anzahl der zu besprühenden Dreiecke zwischen diesen Grenzen liegt. Um die Genauigkeit zu erhöhen, kommt dem Remeshing-Verfahren große Bedeutung zu, da eine größere Anzahl kleinerer Dreiecke erforderlich ist, die zu einer präziseren Lösung beitragen können.

LITERATURVERZEICHNIS

- [1] M. Botsch, L. Kobbelt, M. Pauly, P. Alliez und B. Levy, Polygon Mesh Processing, 2010.
- [2] R. Li, P. Ming, S. Zhiyuan und Z. Yang, „An Arbitrary-Order Discontinuous Galerkin Method with One Unknown Per Element,“ Journal of Scientific Computing, 2019.
- [3] P. Alliez, D. Cohen-Steiner, O. Devillers, B. Lévy und M. Desbrun, „Anisotropic Polygonal Remeshing,“ INRIA, 2003.
- [4] M. Zilske, „Adaptives Remeshing von nicht mannigfaltigen Dreiecksnetzen,“ Freie Universität Berlin, Berlin, 2007.

Photogrammetrische Bestimmung von Kopfbewegungen mittels Facial Landmarks – lessons learned

Mirko Riedel (deveritec), Danilo Schneider (HTW Dresden), Anne Bienert (TU Dresden)

deveritec GmbH
Bamberger Straße 1, 01187 Dresden
eMail: mirko.riedel@deveritec.com
URL: <http://www.deveritec.de>

Abstract. Die moderne Strahlentherapie ist eine hochpräzise Therapieform, die meist über mehrere Wochen bei onkologischen Patienten durchgeführt wird. Bei der Lagerung von Kopf-Tumor-Patienten kommen individuell angefertigte Masken und Kopfstützen zur Anwendung, um reproduzierbare Patientenpositionierung zu garantieren und damit die Präzision der Therapie zu erhöhen. Die tägliche Nutzung dieser Hilfsmittel ist für einige Patienten belastend. Die Umsetzung einer hilfsmittelreduzierten sowie einer (halb-)aufrechten Behandlung lässt sich mit Hilfe von Kamerasystemen realisieren. Über Methoden der Künstlichen Intelligenz (KI) kann eine Merkmalsextraktion von individuellen Gesichts- oder Körperbereichen stattfinden. Eine optische Überwachung der Patienten mit mehreren im Behandlungsraum angebrachten Kameras ermöglicht eine hilfsmittelreduzierte Patientenpositionierung, sowie eine Bewegungskontrolle während der Therapie. In dem Beitrag wird u.a. die Softwarebibliothek Google MediaPipe zur Ableitung von facial landmarks und der daraus abgeleiteten Pose untersucht. Die Ergebnisse der Posenbestimmung liegen unter < 1 mm und zeigen, dass erklärbare KI-Methoden zur Kopfposenbestimmung, eine Überprüfung der Pose sowie der detektierten facial landmarks im Objektraum erfordern, um die Zuverlässigkeit und Reproduzierbarkeit sicherzustellen.

1 DANKSAGUNG

Die Inhalte dieses Beitrages wurden durch das Else Kröner-Fresenius-Zentrum für Digitale Gesundheit (TU-Dresden) als ein „EKfZ – Innovation Projects“ mit dem Namen „Increasing the precision of patient positioning in radiotherapy using positioning aids“ (imPRESSing) gefördert. Das Ziel des Förderprogramms ist dabei technologische Voruntersuchungen und deren Bewertung hinsichtlich der prinzipiellen Einsatzfähigkeit (proof of concept) zur Übertragung innovativer Ideen in Anwendungen im Bereich der Medizintechnik durchführen zu können. Konkret handelt es sich hier, um die Anwendung der Künstlichen Intelligenz (KI)-basierten Detektion von facial landmarks, für die Positionsüberwachung von Patienten in der Präzisionsstrahlentherapie. Die Autoren danken den Reviewern für ihre wertvollen Kommentare.

2 EINLEITUNG UND MOTIVATION

Seit mehr als zwei Jahrzehnten ist die Posendetektion von Personen in Bildern bzw. Videosequenzen (u. a. Moeslund & Granum, 2001; Moeslund et al., 2006) ein Thema in der computer vision. Dies zeigt sich in einer hohen Anzahl unterschiedlicher KI-basierter Gesichts-, Posen- und Bewegungserkennungsmethoden in einzelnen Bildern, was einem hohen Entwicklungstempo geschuldet ist. Beispielhaft sind in Abb. 1 die Bewegungs- und Gesichtsdetektion der Google MediaPipe Bibliothek zusammengestellt. Diese Detektoren sind für den Einsatz auf mobilen Endgeräten optimiert und können in „Echtzeit“, d. h. mit bis zu 20 Hz auf Mittelklasse-Smartphones laufen. Eingesetzt werden sie beispielsweise zum Overlay von Filtern in Videos oder zur Übertragung von Gesichtsausdrücken und Mimik auf Avatare oder Charaktere in Animationsfilmen. Dieser Bereich unterliegt dabei einer recht geringen Regulatorik.

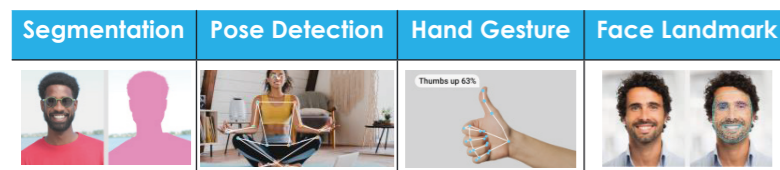
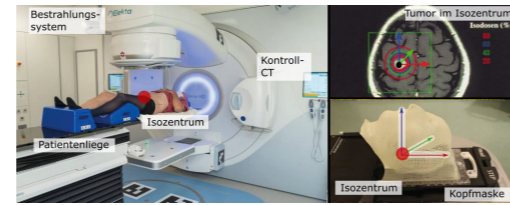


Abb. 1. Beispielhafte Zusammenstellung von „low code“ Beispielen der Bibliothek „Mediapipe“ von Google [Mediapipe, 2023a]

Die Detektion distinktiver Körper- bzw. Gesichtsmerkmale (facial landmarks) mit KI Methoden ist Stand der Technik in medizinischen Anwendungen (Vu et al., 2022). Hammadi et al. (2022) analysierten zehn Algorithmen zur Detektion von Gesichtsmerkmalen (Veröffentlichungszeitraum 2009 bis 2020) und untersuchten die Eignung zum Einsatz in klinischen Anwendungen, wie die Kopf-Posenbestimmung (Hammadi et al., 2022) sowie die Emotionserfassung (Dinculescu et al., 2019; Ko, 2018; Raj & Kumar 2020). Raj & Kumar (2020) extrahierten 68 facial landmarks, aus denen emotionsspezifische Vektoren zum Training einer Support Vector Machine abgeleitet wurden. Das Klassifikationsergebnis zeigt eine Detektionsrate von durchschnittlich 89 % der sieben untersuchten Emotionen.

Der Medizinsektor ist an einer stetigen Entwicklung und Verbesserung neuer Methoden interessiert, was u. a. frühere Diagnosen und bessere Therapien einschließt. Im Bereich der Krebstherapie wird eine schnellere, effizientere und kostengünstigere Krebsbehandlung angestrebt, bei der das Patientenwohl und -sicherheit einen sehr hohen Stellenwert einnehmen. Insgesamt ist im Medizinsektor das Entwicklungstempo recht gering einzustufen, was u. a. auch mit der Zulassung nach der EU-Medizinprodukteverordnung (MDR) sowie allgemein mit dem sehr hohen Niveau hinsichtlich Spezifität und Sensitivität der Methoden zusammenhängt.

Das Niveau der Schutzmaßnahmen, während der Strahlentherapie von Tumorkranken ist sehr hoch. Bei Kopf-Tumor-Patienten wird der Kopf mechanisch fixiert, um sicherzustellen das möglichst nur im Tumor die Strahlendosis zum Absterben des Gewebes führt. Es ist offensichtlich, dass im Bereich der Präzisionsstrahlentherapie, d. h. zur lokalen Behandlung von Tumoren mittels ionisierter Strahlung oder Teilchenstrahlung, durch das optische Tracking (Verfolgung von Bewegungen oder räumlichen Veränderungen über die Zeit) das Patientenwohl entscheidend beeinflusst wird und eine hilfsmittelfreie Patientenpositionierung angestrebt wird. Über das Detektieren von facial landmarks wird die Pose berechnet und mittels kontinuierlichen Tracking deren Änderung während der Bestrahlung bestimmt.



Beispielhafte Positionierung eines Patienten in einem Linearbeschleuniger im Rahmen der Strahlentherapie [links aus Debus et al, 2017]; Maske zur Kopffixierung bei der Bestrahlung eines Hirn-Tumors [rechts aus Schlegel et al., 2018]

Die Kernfragen des Projektes lassen sich wie folgt zusammenfassen:

- 1) Ist es möglich die Methoden zur Detektion von *facial landmarks* hinsichtlich ihrer prinzipiellen Einsatzfähigkeit in Medizinprodukten zu qualifizieren?
- 2) Können Genauigkeiten von unter 1 mm mit einer für Medizinprodukte erwarteten Zuverlässigkeit von einem derartigen Trackingsystem erwartet werden?

Dabei liegt der Fokus dieser Untersuchung in der Genauigkeit und Zuverlässigkeit der Methoden. Die Fragen der Zulassung nach MDR werden nur in dem Maße betrachtet, dass nach prinzipiellen Ausschlusskriterien gesucht aber keine konkrete Bewertung der notwendigen Maßnahmen vorgenommen werden.

3 PROZESSKETTE DES TRACKINGS

3.1 PROZESSKETTE DES TRACKINGS

Die Prozesskette sieht einen Behandlungsprozess vor, der durch das optische Tracking mit mehreren Kameras ergänzt wird. Dabei wird nach der korrekten Positionierung des Patienten im Bestrahler optisch die Pose des Kopfes erfasst und anschließend für das Tracking nur die Posenänderung betrachtet. Wenn die Posenänderung einen definierten Schwellwert überschreitet und die Bestrahlung noch nicht abgeschlossen ist, wird die Behandlung unterbrochen und kann nach einer erneuten Positionierung weitergeführt werden, bis sie abgeschlossen ist (Abb. 3). Diese Herangehensweise ermöglicht eine Behandlung ohne enganliegende Gesichtsmaske. Kernpunkt beim Tracking im 2D-Bild ist die Extraktion der Bildkoordinaten der facial landmarks des zugrundeliegenden Gesichtsmodells. Da sich diese, je nach Erkennungsmethode sehr stark unterscheiden, überführen wir diese in ein charakteristisches Gesichtsmodell auf Basis

von Bezugspunkten (Keypoints). Diese lassen sich unabhängig der eingesetzten Erkennungsmethode mittels räumlichen Vorwärtsschnitt in den 3D-Raum überführen. Die räumliche Anordnung der Keypoints ist dabei für jeden Patienten individuell und werden in ortsfeste und bewegliche Punkte unterteilt. Die ortsfesten Keypoints (z. B. Augen) werden zur Posenbestimmung verwendet, während die veränderlichen Keypoints (z. B. Kinn) die Mimik repräsentieren und für die Posenbestimmung weniger geeignet sind.

Die patientenspezifische eindeutige Anordnung der Keypoints wird als 3DKopfsignatur bezeichnet. Aus den patientenspezifischen 3D-Koordinaten der Keypoints des allgemeinen Kopfmodells wird zu Beginn des Trackings eine „Patientensignatur“ abgeleitet und als Patientendatensatz gespeichert (Abb. 3). Dabei wird momentan angenommen, dass der Tumor sich im Isozentrum der Anlage befindet, was aktuell vom behandelnden Arzt mit Hilfe des Kontroll-CTs überprüft wird. In den folgenden Schritten wird nur noch die Positions- und Orientierungsänderung, also die Pose der Signatur relativ zum Startzeitpunkt bestimmt. Daraus werden die lokalen Verschiebungen der Tumorposition berechnet und dargestellt.

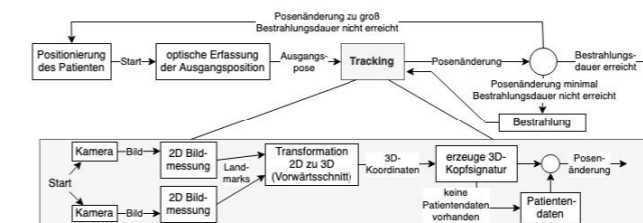


Abb. 3. Prozessablauf der optisch überwachten Präzisionsstrahlentherapie

3.2 VERSUCHSAUFBAU

Zum Test der Applikation wurde ein Demonstrator realisiert (Riedel et al., 2022), der mittels eines definiert positionierbaren Puppenkopfes eine Bewegung erzeugt, die mit Trackingmethoden messbar ist (Abb. 4). Die Auswertesoftware wurde so gestaltet, dass zum einen die Positionierung zum Hintergrundpattern (später Patientenliege) mit verschiedenen Messmarkentypen durchgeführt werden kann (z. B. ArUcos in Abb. 4) und zum anderen unterschiedliche KI-Operatoren zur Erfassung von facial landmarks (z. B. Google MediaPipe, LBF, MTCNN) verwendet werden können. Abbildung 4 (rechts) zeigt die verschiedenen Kalibrierobjekte die in Riedel et al. (2022) untersucht wurden.



Abb. 4. Demonstrator aufgenommen von 3 Kamerapositionen (links); positionierbarer Puppenkopf und Hintergrundpattern (Mitte); Kalibrierobjekte zur Kalibrierung des Demonstrators (rechts) [Riedel et al., 2022]

3.3 KI-BASIERTE BILDMESSUNG

Methoden die mit KI-Systemen arbeiten sind undurchsichtig und es ist für den Nutzer nur sehr schwer bis gar nicht nachvollziehbar warum der Algorithmus diese Entscheidung getroffen hat. Prinzipiell handelt es sich dabei um Methoden mit einem Blackbox-Verhalten. Mit einer „erklärbaren“ KI kommt es zu einer Steigerung der Nachvollziehbarkeit beim Nutzer. Dabei meint der Begriff der „erklärbaren“ KI:

„Erklärbare KI-Systeme liefern eine Beschreibung der Ergebnisfindung, die für Menschen interpretierbar ist. Das heißt, die Gründe für eine Entscheidung lassen sich explizit beschreiben und können den Adressat:innen eines Systems verständlich vermittelt werden. Erklärbarkeit ist dabei nicht mit Transparenz gleichzusetzen: Zum einen können auch Erklärungen generiert werden, ohne das Innenleben eines Algorithmus nachzuzeichnen, zum anderen kann ein transparentes System zu komplex sein, um es mit vertretbarem Aufwand nachvollziehen zu können.“ (Kompetenzzentrum Öffentliche IT, 2023)“

Bei der erklärbaren KI werden heatmaps verwendet, um entscheidungsrelevante Bereiche darzustellen und zu beurteilen. Dies ermöglicht eine visuelle Kontrolle, auf Basis welcher Merkmale im Bild die KI ihre Entscheidung trifft.

Es wurden verschiedene Gesichts- und facial landmark-Detektoren implementiert und experimentell getestet:

- Cascade Classifier (Viola & Jones, 2001)
- DLIB HOG und SVN (Dala & Triggs, 2005)
- DLIB CNN (Jadhav et al., 2021)
- OpenCV DNN (Nagrath et al., 2016)
- MTCNN (Zang et al., 2016)
- DLIB – facial Landmarks (Kazemi & Sullivan, 2014)
- OpenCV FacemarksLBF (Ren et al., 20014)
- MediaPipe (Bazarevsky et al., 2019)

Für weiterführende Untersuchungen wird auf Kramp (2022) verwiesen.

3.4 3D-TRANSFORMATION DER FACIAL LANDMARKS

Um die Ergebnisse der KI-Bildmessung prüfbar zu gestalten, wird ein Mehrkamerasystem mit einer mehrfach redundanten Gesichtserkennung verwendet, um fehlerhafte Ergebnisse in einer Plausibilitätskontrolle zu detektieren. Grundlage dafür ist eine Systemkalibrierung mit einem ausreichend hohem Genauigkeitspotenzial (vgl. Riedel et al., 2022). Es wurde eine Kopfbeschreibung definiert, in der die einzelnen facial landmarks in operatorunabhängige Keypoints konvertiert werden, um Ergebnisse unterschiedlicher Modelle in das gleiche Kopfmodell zu überführen.

3.5 TEST UND VERIFIKATION

Zur Verifikation der Messung wurde der Puppenkopf auf einen Trossen Robotics Viper 300 Roboter (Trossen, 2023) mit einer theoretischen Tragkraft von 750 Gramm und einer angegebenen Wiederholgenauigkeit von 1 mm montiert und kontrolliert einer Bewegung unterzogen. Unter realen Bedingungen dürfte die erreichbare Genauigkeit jedoch schlechter sein. Trotzdem kann damit eine definierte Pose über die Robotersteuerung angefahren werden, die dann über die Bildmessung in gleichem Maße erfasst werden sollte. Dies macht den Versuchsaufbau in einem Bereich von unter 5 mm bis max. 1 mm überprüfbar.

4 ERGEBNISSE

4.1 KI-BASIERTE BILDMESSUNG

Nicht für alle im Beitrag verwendeten Operatoren sind heatmaps zur Visualisierung der Entscheidungsfindung verfügbar, weshalb auf die Bibliothek 3D-Alignment (Deng et al., 2018) zurückgegriffen wurde (Lime-Repräsentation, Abb. 5, rechts). Diese lässt sich den verwendeten Operatoren (MTCNN, LBF, MediaPipe) gut

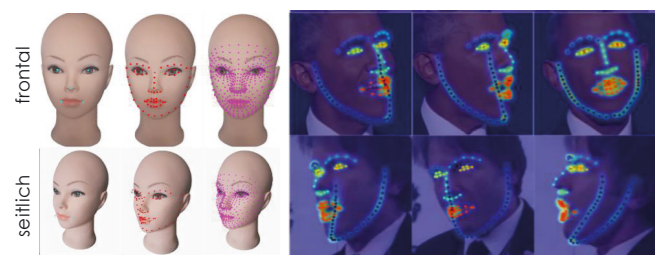


Abb. 5. Beispielfähige Auswahl an detektierten facial landmarks (CNN, DNN, Mediapipe) (links); Darstellung der LIME-Repräsentationen der Landmarks der Bibliothek 3D-Alignment [Deng et al., 2018]

gegenüberstellen (Abb. 5, links). Dabei sind die Punkte hinsichtlich ihrer Zuverlässigkeit von Rot bis Blau (hoher bis geringer Einfluss) dargestellt. Die ausgewählten landmark Detektoren wurden hinsichtlich ihrer Wiederholgenauigkeit untersucht. Dabei wurden 2 markante Keypoints (Nasenspitze und innere linke Augenpunkt) über 100 Frames in einer Videosequenz (1280 x 720 Pixel) beobachtet

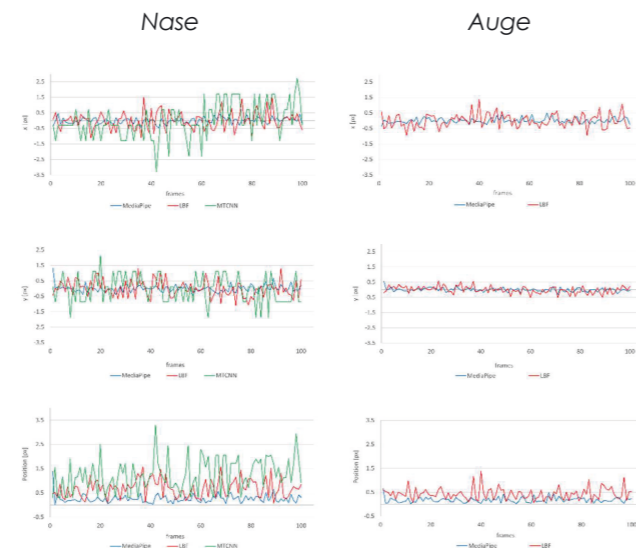


Abb. 6. Koordinaten- und Keypointabweichungen von Nasenspitze (links) und linker innerer Augenpunkt (rechts) ermittelt mit MediaPipe (blau), LBF (rot) und MTCNN (grün) über 100 Frames

(Abb. 6). Die MediaPipe Bibliothek zeigte die geringsten Koordinatenabweichungen über 100 Frames. Weitere Untersuchungen konzentrieren sich auf die MediaPipe Bibliothek, da sie zum einen die besten Ergebnisse erzielt hat und zum anderen die MTCNN Bibliothek mit nur 5 extrahierten landmarks (inkl. bewegbarer Irismittelpunkte) für die Posenbestimmung zu wenig überbestimmt und ungeeignet ist.

4.2 MEDIAPIPE

Wie schon die oben aufgeführte Definition ausdrückt, ist die „Erklärbarkeit“ der Entscheidungsfindung in Hinblick auf Genauigkeit und Sicherheit für den Einsatz im medizinischen Tracking nicht hilfreich. Mit dem Wissen, wieso die KI diese Entscheidungen getroffen hat, ist eine Überprüfung, ob diese Entscheidungen richtig und ob das Ergebnis im 3D-Raum auch genau ist, nicht möglich. Die untersuchten Operatoren liefern kein Qualitätskriterium an den Nutzer zurück. Bei einigen Operatoren kann ein Confidence-Wert C im Funktionsaufruf eingestellt werden. Dies bezweckt eine Filterung und nur „sichere“ Ergebnisse werden ausgegeben. Was dabei als sicher eingestuft wird, ist nicht ersichtlich. Im Allgemeinen sind

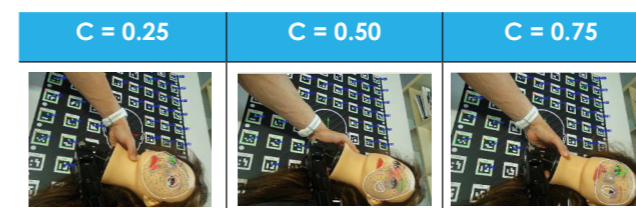


Abb. 7. Beispiel für provozierte fehlerhafte Bildmessungen bei unterschiedlich eingestellten Confidence-Werten im Bereich von $0.25 \leq C \leq 0.75$

die Methoden der untersuchten Operatoren klassische Blackbox-Algorithmen. Mit der Bibliothek MediaPipe (MediaPipe, 2023b) lassen sich dabei die Grenzen der Confidence-Werte recht einfach austesten (Abb. 7).

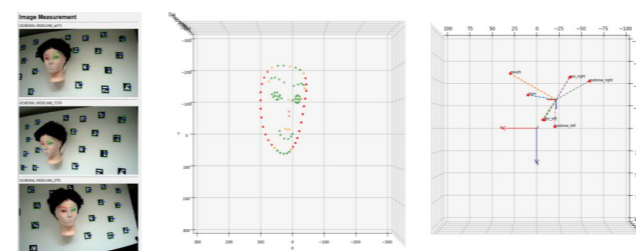


Abb. 8. Transformation der facial landmarks (links) vollständig (Mitte) bzw. mittels keypoints (rechts) in den 3D-Raum

4.3 3D-TRANSFORMATION DER FACIAL LANDMARKS

Erwartungsgemäß weisen die von der KI schlecht erkennbaren Konturpunkte eine signifikant höhere Unsicherheit auf als die gut erkennbaren, texturreichen Keypoints wie Nase, Auge und ggf. auch Mund (Abb.8). Einerseits korrespondiert dieses Verhalten mit den Aussagen zur „Erklärbarkeit“ der Erfassung der Merkmale, andererseits lässt dieses Vorgehen auch teilweise eine Trennung von Position des Kopfes und Mimik des Patienten zu.

Feature	Sigma0	X	SX	Y	SY	Z	SZ
nose	5.54	-6.14	1.87	-32.20	2.18	204.43	2.28
mouth	4.26	-15.87	1.45	-55.72	1.62	180.82	1.89
eye_left	5.01	21.89	1.85	1.62	2.28	168.30	2.41
eye_right	4.54	-33.42	1.59	1.30	2.09	170.01	2.22
iris_left	3.90	-35.87	1.36	-0.67	1.78	172.67	1.89
iris_right	7.62	21.26	2.80	2.84	3.48	170.08	3.65
eyebrow_left	7.09	25.43	2.63	20.89	3.37	175.51	3.38
eyebrow_right	6.08	-37.37	2.14	21.12	2.94	176.08	2.96

Tab. 1. Beispielfähige Messung der acht keypoints (Feature) mit Unsicherheit des Vorwärtsschnittes (Sigma0; in Pixel), den 3D-Koordinaten (X, Y, Z; in mm) sowie den zugehörigen Unsicherheiten (SX, SY, SZ; in mm)

4.4 TEST UND VERIFIKATION

Zur Verifikation der Methode werden neben den Koordinaten der Keypoints auch die Unsicherheiten, welche aus dem räumlichen Vorwärtsschnitt zur Berechnung der 3D Koordinaten gewonnen werden, mit ausgegeben (Tab. 1). Dabei zeigt sich, dass mit den verwendeten Erkennungsoperator (MediaPipe) und den eingestellten Parametern, ein Bereich der Unsicherheiten der Ausgleichung von 3.9 Pixeln bis 7.6 Pixeln im Bildraum erreicht wird. Übertragen in den Objektraum, entspricht das einem Bereich der Unsicherheit der Keypoints von 2.9 mm bis 5.8 mm.

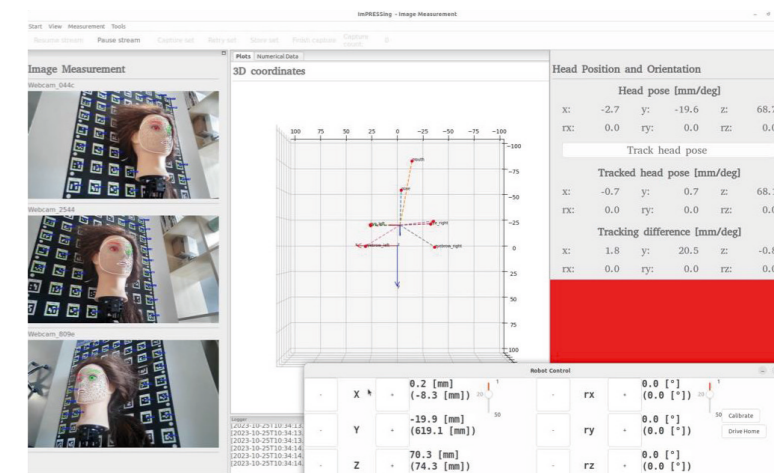


Abb. 9. Beispielfähiger Screenshot der GUIs zur Erfassung der Patientenpose (imPRESSing - image measurement) überlagert mit der Bewegungsvorgabe des Roboters (Robot Control)

In die Berechnung der Unsicherheit der Keypoints gehen prinzipiell alle Bildmessungen aller Kameras ein. Die stochastische Unsicherheit der Bildmessungen zeigt sich in den Keypoints im Sigma0 (Pixel) sowie in SX, SY und SZ (mm, Tab.1). Diese Unsicherheit darf jedoch nicht mit der Genauigkeit der Koordinaten (X, Y, Z), d.h. der absoluten Position des Bildpunktes im Bild, gleichgesetzt werden (vgl. Luhmann, 2018).

Im nachgelagerten Schritt werden aus den Keypoints die Lage und Orientierung der Kopfpose sowie die Kopfposenänderung bestimmt. Es geht nur noch die charakteristische Anordnung der Keypoints (hinterlegt in den Patientendaten) in die Berechnung ein. Dabei zeigt sich, dass die Anordnung der Keypoints zueinander konstant ist und die Änderung der Pose von Zeitpunkt t_0 zu t_1 die Lage um maximal 1.44 mm bei einer inneren Unsicherheit von 0.52 mm variiert (Tab. 2). Dies spricht für die Annahme, dass die Keypoints zwar verrauscht sind, jedoch relativ genau bestimmt werden können.

	Position [mm]		Rotation [rad]	
X	1.44	±0.52	0.15	±0.0012
Y	0.46	±0.47	-0.45	±0.0014
Z	0.38	±0.13	-0.11	±0.0010

Tab. 2. Beispielhafte Unsicherheiten einer aus acht Keypoints abgeleiteten Kopfposenänderung

Die aus acht Keypoints abgeleitete Kopfpose des Patienten (Im nachgelagerten Schritt werden aus den Keypoints die Lage und Orientierung der Kopfpose sowie die Kopfposenänderung bestimmt. Es geht nur noch die charakteristische Anordnung der Keypoints (hinterlegt in den Patientendaten) in die Berechnung ein. Dabei zeigt sich, dass die Anordnung der Keypoints zueinander konstant ist und die Änderung der Pose von Zeitpunkt t_0 zu t_1 die Lage um maximal 1.44 mm bei einer inneren Unsicherheit von 0.52 mm variiert (Tab. 2). Dies spricht für die Annahme, dass die Keypoints zwar verrauscht sind, jedoch relativ genau bestimmt werden können.

Tabelle 2 zeigt eine geringe Unsicherheit von unter 1 mm (Position) bzw. 0.1° (Orientierung). Das entspricht prinzipiell bereits dem geforderten Genauigkeitsniveau der Anwendung.

5 DISKUSSION

Auffällig ist, dass vor allem Gesichtsbereiche mit einer gut ausgeprägten optischen Textur zur facial landmark Detektion herangezogen werden. Dies sind vor allem Mund- und Augenpartie sowie Bereiche der Nase. Die Kinnpartie, die sich im Bild durch geringe Helligkeits- oder Farbgradienten auszeichnet, wird auch hier eher gering bewertet.

Die Untersuchungen zur Variation der Confidence-Werte verdeutlichen, dass unabhängig von der eingestellten Vertrauensschwelle, Fehlmessungen auftreten. In den ausgewählten Bereichen trifft die KI dabei die falsche Zuordnung von Mund und Auge und erzeugt dennoch stabile Ergebnisse. Diese Fehlinterpretation ist für die Methode selbst nicht erkennbar. Fazit ist dementsprechend, dass die Blackbox nicht auflösbar ist und eine Prüfbarkeit außerhalb der Bildmessung realisiert werden muss.

Der Verifikationstest bestätigt eine prinzipielle Fähigkeit der Methoden für die am Anfang aufgeführte Problem-

stellung. Bei der Einordnung der Unsicherheiten muss berücksichtigt werden, dass für die Untersuchungen einfachen Webcams eingesetzt wurden, deren Bilder mit nicht für den Anwendungsfall (konvergente Bildaufnahme) optimierte Operatoren zur Detektion der facial landmarks ausgewertet werden. Zudem wurde ein Roboterarm mit einer Wiederholgenauigkeit von 1 mm eingesetzt. Es wäre denkbar, dass der Roboterarm sowie die KI-Methoden beide einen fehlerhaften Positionswert ermitteln, die zu einer zu optimistischen Positionsgenauigkeit führt. Um dies auszuschließen, könnte zum Einen der Roboterarm zusätzlich mit Messmarken photogrammetrisch getrackt werden und zum Anderen ein Roboterarm höherer Genauigkeit eingesetzt werden. Der exakte Ground-Truth-Wert der Kopfpose ist für weiterführende Tests und Entwicklungen des Systems unabdingbar.

Zusammengefasst können, basierend auf den vorgestellten Ergebnissen, die zwei gestellten Kernfragen einfach beantwortet werden:

1. Der vorgestellte Aufbau kann die Genauigkeit und Zuverlässigkeit evaluieren.
2. Trotz Verwendung von einfachen Webcams und einem Roboter mit einer mittleren Genauigkeit zeigten sich zufriedenstellende Ergebnisse. Mit einem kalibrierten Mehrkamerasystem zur redundanten Bildaufnahme kann die Pose des Kopfes (< 1 mm) zuverlässig erfasst werden.

6 FAZIT

Es ist beachtlich, was mit den verfügbaren Operatoren („low code“ Methoden) realisiert werden kann, solange man die Plausibilität der Ergebnisse hinterfragt. Die wichtigsten Erkenntnisse der Studie „lessons learned“ sind:

- Der bestehende Messaufbau, Bilderkennung, Transformation von Bild- in Objektraum und Ableitung der Pose muss als geschlossene Prozesskette betrachtet werden (Abb. 3).
- Eine Bewertung der Ergebnisse der KI-Bilderkennung aus sich heraus ist nicht möglich. Auch die Methoden zur Herstellung der Erklärbarkeit der Erkennung (heatmaps) sind für eine Bewertung von Genauigkeit und Zuverlässigkeit nicht ausreichend (Abb. 5).
- Die erklärbare KI-Erkennung ist eine Blackbox von der auszugehen ist, dass zuverlässig falsche Ergebnisse produziert werden.

- Die Qualität der Ergebnisse ist von den Eingangsdaten abhängig, auch wenn das KI-Modell geeignet ist.
- Hinsichtlich der Kopfposenbestimmung mit erklärba- ren KI-Methoden, muss eine Überprüfung der Pose sowie der detektierten facial landmarks im Objekt- raum erfolgen, um die Zuverlässigkeit und Reprodu- zierbarkeit sicherzustellen.
- Es ist möglich ein System zu realisieren, von dem ein medizinischer Einsatz nicht prinzipiell ausgeschlossen ist. Eine Mehrfachmessung der facial landmarks, ggf. auch durch Kombination von unterschiedlichen KI-Operatoren, mit anschließender Fusion führt zu einer Plausibilitätsprüfung.

Oder, um es in einem Satz zusammenzufassen:

Die Erkennung der facial landmarks mit KI wird innerhalb dieses Systems zwar nicht „erklärbar“ aber im weiteren Sinne „überprüfbar“.

LITERATURVERZEICHNIS

Bazarevsky, V., Kartynnik, Y., Vakunov, A., Raveendran, K., Grundmann, M., 2019: BlazeFace: Submillisecond Neural Face Detection on Mobile GPUs. CVPR Workshop on Computer Vision for Augmented and Virtual Reality, Long Beach, CA, USA, 2019

Dala, N., Triggs, B., 2005: Histograms of oriented gradients for human detection. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 886-89.

Debus, P., Bartenstein, J., Reiser, M., Kuhn, F.-P., 2017: Duale Reihe, Radiologie. 4. vollständig überarbeitete Auflage. Thieme Stuttgart.
Deng, J. Y., Zhou, S., Cheng, Z., Zaferiou, S., 2018: Cascade Multi-View Hourglass Model for Robust 3D Face Alignment. 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 399-403.

Dinculescu, A., Baltoiu, A., Strungaru, C., Petrescu, L., Vizitiu, C., Mandu, A., Talpes, N., Valeanu, V., 2019: Automatic Identification of Anthropological Face Landmarks for Emotion Detection. 9th International Conference on Recent Advances in Space Technologies (RAST), Istanbul, Turkey, 585-590.

Hammadi, Y., Grondin, F., Ferland, F., Lebel, K., 2022: Evaluation of Various State of the Art Head Pose Estimation Algorithms for Clinical Scenarios. Sensors, 22(18), 6850.

Jadhav, A., Lone, S., Matey, S., Madamwar, T., Jakhete, S., 2021: Survey on Face Detection Algorithms. International Journal of Innovative Science and Research Technology, 291-297.

Kazemi, V., & Sullivan, J., 2014: One millisecond face alignment with an ensemble of regression trees. IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, 1867-1874.

Ko, B., 2018: A Brief Review of Facial Emotion Recognition Based on Visual Information. Sensors, 18(2), 401.

Kompetenzzentrum Öffentliche IT, 2023: <https://www.oeffentliche-it.de/-/erklarbare-ki>, Stand: 25.10.2023

Kramp, A., 2022: Entwicklung und Validierung eines Konzepts zur photogrammetrischen Posen-Bestimmung von Köpfen in der Strahlentherapie auf Basis eines Vier-Kamera-Systems. Masterarbeit am Institut für Photogrammetrie und Fernerkundung, TU Dresden (unveröffentlicht).

Luhmann, T., 2018: Nahbereichsphotogrammetrie. Grundlagen – Methoden – Beispiele. 4. neu bearbeitete Auflage, Wichmann Verlag.

MediaPipe, 2023a: <https://developers.google.com/mediapipe/solutions/examples>, Stand 25.10.2023

MediaPipe, 2023b: https://mediapipe-studio.webapps.google.com/studio/demo/face_landmarker, Stand 25.10.2023

Moeslund, T. B. & Granum, E., 2001: A Survey of Computer Vision-Based Human Motion Capture. Computer Vision and Image Understanding, 81(3), 231-268.

Moeslund, T. B., Hilton, A., Krüger, V., 2006: A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding, 104(2-3), 90-126.

Nagrath, P., Jain, R., Madan, A., Arora, R., Kataria, P., Hemanth, J., 2016: SSDMN2: A real time DNNbased face mask detection system using single shot multibox detector and MobileNetV2. Sustainable Cities and Society 66(6789):102692.

Raj S., R., Pratiba, D., Kumar P, R., 2020: Facial Expression Recognition using Facial Landmarks: A Novel Approach. Advances in Science, Technology and Engineering Systems Journal, 5(5), 24-28.

Ren, S., Cao, X., Wei, Y., Sun, J., 2014: Face Alignment at 3000 FPS via Regressing Local Binary Features. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, 1685-1692.

Riedel, M., Schneider, D., Bienert, A., 2022: Photogrammetrische Bestimmung von Kopfbewegungen mittels facial Landmarks – proof of concept -. Tagungsband des 24. Anwendungsbezogener Workshop zur Erfassung, Modellierung, Verarbeitung und Auswertung von 3D-Daten, Berlin, 2022, ISBN 978-3-942709-29-3

Schlegel, W., Karger, C. P., Jäkel, O., 2018: Medizinische Physik Grundlagen – Bildgebung – Therapie – Technik. Springer Berlin Heidelberg.

Trossen, 2023: <https://www.trossenrobotics.com/viperx-300-robot-arm-6dof.aspx>, Stand 25.10.2023

Viola, P. & Jones, M., 2001: Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, I-I.

Vu, N. H., Trieu, N. M., Anh Tuan, H. N., Khoa, T. D., Thinh, N. T., 2022: Review: Facial Anthropometric, Landmark Extraction, and Nasal Reconstruction Technology. Applied Sciences, 12(19), 9548.

Zhang, K., Zhang, Z., Li, Z., Gao, Y., 2016: Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. IEEE Signal Processing Letters 23.10, 1499-1503.

Identifikation klinisch interpretierbarer Feature-Vektoren für die Analyse von Torso Surface-Topographie Daten

Stephan Rothstock (GFai, Berlin), José María González-Ruiz (University of Alberta), Hans-Rudolf Weiss (Koob ScolioTech GmbH, Neu-Bamberg), Deborah Turnbull (Physiotherapist, ScoliosisUK), Daniel Krueger (GFai, Berlin)

Gesellschaft zur Förderung angewandter Informatik e.V.
Volmerstrasse 3, 12489 Berlin
eMail: rothstock@gfai.de
URL: <http://www.gfai.de>

Abstract. Surface Topographie-Daten des menschlichen Oberkörpers werden üblicherweise im Rahmen von Skoliose Screenings mit Hilfe von 3D-Scannern erhoben. Die daraus resultierenden Punktwolken bzw. Oberflächen-Triangulationen stellen aufgrund der Menge an Datenpunkten ($n > 10.000$) verglichen mit der zur Verfügung stehenden Anzahl an Trainingsdaten ($m < 100$) eine Herausforderung für datengetriebene (herkömmliche maschinelle) Lernverfahren dar. Hier können semantische Informationen (Expertenwissen) von anatomischen bzw. klinischen Zusammenhängen helfen, leistungsfähige und klinisch interpretierbare Feature-Vektoren zu identifizieren, welche zudem die Dimension der Daten stark reduzieren ($n < 10$). Die Arbeit kann demnach in das Forschungsfeld der 'Explainable & Interpretable AI' eingeordnet werden.

1 EINLEITUNG

Die Charakterisierung des Phänotyps, also der Oberkörperform, von Skoliose-Patienten anhand von 3D-Scans ist neben den röntgenbasierten Diagnosetools eine der wichtigsten Analysemethoden um den Schweregrad von Skoliose zu bestimmen oder eine Behandlungsempfehlung mit einer Korsett-Therapie abzuleiten (Applebaum et al. 2021). Die durch 3D-Scan Techniken erhobenen Daten bestehen üblicherweise aus mehr als 100.000 Datenpunkten bzw. Dreiecken, welche die Torso-Oberfläche diskretisieren. Demgegenüber stehen in den meisten Studien weniger als 100 Patienten für das Training von Klassifizierungsalgorithmen zur Verfügung. Das Problem ist in der Literatur als $n \gg m$ Fall bekannt und resultiert bei rein datengetriebenen Verfahren (machine learning) in einem 'Over fitting' und einer reduzierten Generalisierungsleistung (Lausser et al. 2022). Bei der Anwendung auf neue Datensätze kann dies zu einer erhöhten falsch-Positivrate führen. Ziel dieser Arbeit ist die Identifikation von interpretierbaren geometrischen Feature-Vektoren, welche durch semantisches Vorwissen über die Anatomie und die klinischen Zusammenhänge bei Skoliose-Patienten abgeleitet werden und gleichzeitig leistungsfähige Klassifizierungen in klinische Patientengruppen erlauben.

2 METHODE

Die Studie wird mit einem Patientekollektiv von 50 Skoliosepatienten durchgeführt und basiert methodisch weitestgehend auf einer bereits veröffentlichten Arbeit (Rothstock et al. 2020). Dabei werden alle Patienten 3D gescannt und gleichzeitig anhand eines Röntgenbildes

durch einen Kliniker klassifiziert (Schweregrad, Cobb-Winkel). Sämtliche weitergehende Analysen in dieser Arbeit basieren auf einem bereits publizierten Workflow, bei dem eine punktbasierte Asymmetriekarte zwischen rechter und linker Körperhälfte durch eine vorausgehende Spiegelung des Torsos an der Sagittalachse erstellt wird (Komeili et al. 2015; Ghaneei et al. 2018). Die Asymmetriekarte entspricht einem Skalarfeld des Patiententorsos (Punktwolke), wobei jedem Punkt die Distanz zu seinem nächsten gespiegelten Nachbarpunkt zugeordnet wird (Abbildung 1).

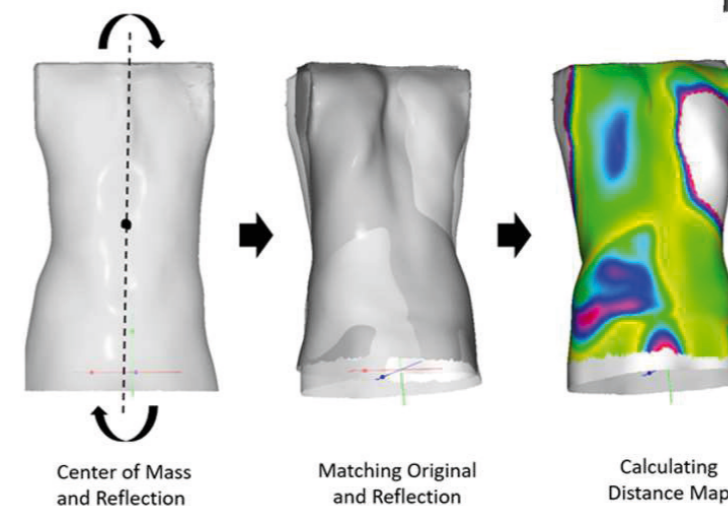


Abb. 1. Darstellung des 3D-Scan basierten Workflows zur Berechnung einer Asymmetriekarte zwischen linker und rechter Torso-Seite.

Für die Erstellung eines aussagekräftigen und klinisch interpretierbaren Feature-Vektors wird semantisches Vorwissen verwendet, welches in Form von Literatur über das Krankheitsbild Skoliose vorliegt. In dieser Arbeit wurden anhand des klinischen Vorwissens fünf Merkmale extrahiert (Abbildung 2).

- Max. Asymmetrie-Distanz (MaxDist)
- Patch-Fläche (PatchArea)
- Lateraler Patch-Abstand (LatDev)
- Surface-Rotation im Patch-Zentrum (SurfRot)
- Form des Asymmetrie-Histogramms (Distance Histogram)

Grundvoraussetzung ist, dass die Merkmale aus der Asymmetriekarte extrahiert werden können und klinisch interpretierbare Eigenschaften aufweisen. Merkmal (a) ist die maximale Asymmetrie-Distanz, welche Auskunft über die lokalen Extrema der Deformation des Patienten gibt. Merkmal (b) ist die Patch-Fläche, also der Teil der Torso-Oberfläche der über einem bestimmten Schwell-

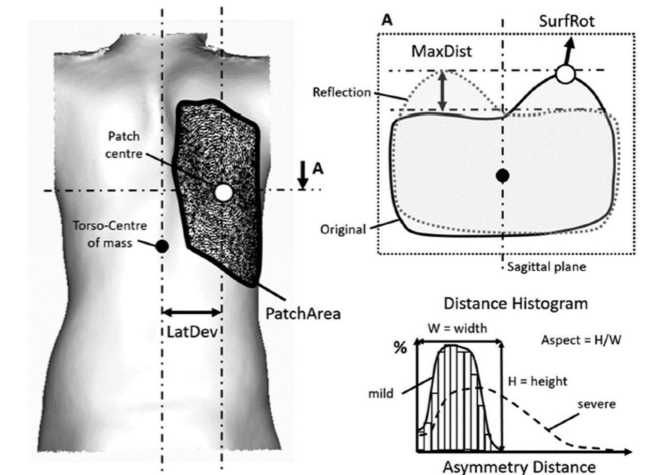


Abb. 2. Darstellung der Herleitung fünf anatomisch und klinisch interpretierbarer Merkmale zur Erstellung eines dimensionslosen Feature-Vektors Π .

wert (> 9 mm) liegt (Ghaneei et al. 2018). Dieser Parameter beschreibt allgemein das Ausmaß (die prozentuale Verbreitung) der asymmetrischen Deformationen. Merkmal (c) beschreibt die laterale Entfernung des Flächenzentrums der größten Deformation (Patches) von der sagittalen Symmetrieachse (Körperhochachse) und damit wie stark die Wirbelsäule nach lateral verkrümmt ist. Merkmal (d) beschreibt die Oberflächenrotation im Zentrum des größten Patches und spiegelt damit die Abweichung von im gesunden Zustand perfekt an der Sagittalebene ausgerichteten Wirbelkörpern ohne Rotation wider. Das letzte Merkmal (e) charakterisiert allgemein das Ausmaß der Asymmetrie des Patienten, indem das Aspekt-Verhältnis (Höhe zu Breite) des normierten Asymmetrie-Distanz-Histogramms berechnet wird. Diese fünf Merkmale werden anschließend durch Normierung (0-1) anhand der Maxima von allen Patienten zu einem dimensionslosen Feature-Vektor zusammengefasst. Zur Verdeutlichung der Leistungsfähigkeit des Feature-Vektors, wird dieser für die in der Skoliose Diagnose wichtige Klassifizierung des Schweregrads (mild oder schwer) angewendet. Als Klassifizierungsverfahren wird die 'k-nearest-neighbor' Suche (k-NN, $k=5$ Nachbarn) durch Berechnung der Euklidischen Distanz d zwischen den Feature-Vektoren eines Trainings- und Testdatensatz eingesetzt (Ghaneei et al. 2019).

$$d = \sqrt{(P_{1Train} - P_{1Test})^2 + \dots + (P_{5Train} - P_{5Test})^2}$$

Dabei wird ein Testdatensatz mit den fünf nächsten Nachbarn des Trainingskollektivs verglichen und dem Label (mild, schwer) zugeordnet, welches am häufigsten vertreten ist ($n > 2$ von 5). Das Training und die Validierung erfolgen mit vorher durch einen Kliniker ausgewerteten Röntgenbildern, in denen anhand des Cobb-Winkels der Schweregrad bestimmt wird. Als Leistungskriterium wird die Sensitivität SE (richtig krank erkannt Rate), Spezifität SP (richtig gesund erkannt Rate) sowie die Genauigkeit ACC (Kombination aus SE und SP) der Klassifizierung ausgewertet.

3 ERGEBNISSE

In Abbildung 3 ist die Berechnung des Feature Vektors P_i bestehend aus fünf Merkmalen getrennt für ein Test- und ein Trainingskollektiv dargestellt.

Test	Thoracic		Lumbar		Thoracic		Lumbar		Thoracic		Lumbar	
	P_1	P_1	P_2	P_2	P_3	P_3	P_4	P_4	P_5	P_5	P_5	P_5
Max Dist	0.455	0.482	0.312	0.432	0.300	0.154	0.370	0.200	0.257	0.392	0.392	0.392
Min	0.176	0.322	0.152	0.162	0.040	0.000	0.080	0.022	0.110	0.118	0.118	0.118
Max	0.927	1.000	0.952	0.710	1.000	0.731	0.819	0.878	1.000	1.000	1.000	1.000
Variance	0.042	0.034	0.046	0.020	0.074	0.046	0.048	0.043	0.043	0.053	0.053	0.053
Train	P_1	P_1	P_2	P_2	P_3	P_3	P_4	P_4	P_5	P_5	P_5	P_5
Median	0.416	0.572	0.352	0.344	0.540	0.231	0.514	0.390	0.222	0.473	0.473	0.473
Min	0.173	0.270	0.000	0.154	0.000	0.000	0.026	0.045	0.067	0.127	0.127	0.127
Max	1.000	0.953	1.000	1.000	0.980	1.000	1.000	1.000	0.796	0.873	0.873	0.873
Variance	0.059	0.025	0.071	0.031	0.120	0.060	0.101	0.058	0.034	0.043	0.043	0.043

Abb. 3. Darstellung des dimensionslosen Feature-Vektors P_i für ein Test ($n=25$) und ein Trainings-Kollektiv ($n=25$) von Patienten.

Die Klassifizierung in eine milde und eine moderat bis schwer erkrankte Behandlungsgruppe ergab eine Genauigkeit (ACC) von thorakal 79 % und lumbal 75 %. Die Ergebnisse für die Sensitivität (SE) und Spezifität (SP) betragen thorakal (SE: 93.3 %, SP: 55.6 %) und lumbal (SE: 88.9 %, SP: 33.3 %). Damit sind die Ergebnisse vergleichbar mit den aktuell besten veröffentlichten internationalen Studien (ACC: 80 %/72 %, SE: 81 %/93 %, SP: 79 %/53 %) jeweils thorakal und lumbal (Ghaneei et al. 2019).

4 DISKUSSION UND AUSBLICK

In der hier vorgestellten Arbeit wurde das generelle Problem vieler Klassifizierungsverfahren beschrieben, dass der Merkmalsraum wesentlich größer ist ($n > 10.000$) als die Menge an zur Verfügung stehenden Trainingsdaten ($m < 100$). Dies kann beim Einsatz von rein datengetriebenen Verfahren (machine learning), welche den

gesamten Merkmalsraum direkt für das Training und die Analyse nutzen, zum Verlust an Generalisierungsleistung und einem 'over-fitting' bei neuen Datensätzen führen (Lausser et al. 2022). Alternative 'machine-learning' Verfahren wie PointNet erstellen ebenfalls niedrig dimensionale Feature Vektoren ($n \sim 1000$) aus einer Menge an zufällig ausgewählten Punkten der Geometrie. Die Feature-Vektoren weisen jedoch keine geometrisch oder klinisch interpretierbaren Zusammenhänge auf und sind rein abstrakter Natur (Qi et al. 2016). Die hier vorgestellte Arbeit verfolgt den Ansatz der Nutzung von semantischem Vorwissen aus anatomischen bzw. klinischen Literaturquellen um einen interpretierbaren geometrischen Feature-Vektor zu erstellen, welcher eine wesentlich kleinere Dimension besitzt ($n=5$) als die Anzahl zur Verfügung stehender Trainingsdatensätze. Ähnliche Ansätze werden bereits in der Klassifizierung von genetischen Profilen im Feld der Bioinformatik erprobt (Lausser et al. 2018). Damit wird das Problem des 'over-fitting' vermieden und es werden weniger Trainingsdatensätze als bei

datengetriebenen Verfahren benötigt.

Der Feature-Vektor konnte erfolgreich für die Klassifizierung von Skoliose-Patienten eingesetzt werden und erzielte international vergleichbare leistungsfähige Ergebnisse für die Sensitivität und Spezifität. Ein großer Vorteil ist die niedrige Dimension ($n=5$) sowie die klinische Interpretierbarkeit der einzelnen Merkmale des Feature-Vektors, welche es Klinikern ermöglicht direkte Rückschlüsse auf die Ausprägung des Krankheitsbildes Skoliose zu ziehen. Die Methode wurde aktuell auch international veröffentlicht (Rothstock et al. 2023) und das Verfahren soll perspektivisch in einer webbasierten Anwendersoftware vermarktet werden.

LITERATURVERZEICHNIS

Applebaum, Ariella; Cho, Woojin; Nessim, Adam; Kim, Keonhee; Tarpada, Sandip P.; Yoon, Seung Ho et al. (2021): Establishing the validity of surface topography for assessment of scoliosis: a prospective study. In: Spine deformity 9 (3), S. 685–689. DOI: 10.1007/s43390-020-00260-9.

Ghaneei, Maliheh; Ekyalimpa, Ronald; Westover, Lindsey; Parent, Eric C.; Adeeb, Samer (2019): Customized k-nearest neighbourhood analysis in the management of adolescent idiopathic scoliosis using 3D markerless asymmetry analysis. In: Computer methods in biomechanics and biomedical engineering 22 (7), S. 696–705. DOI: 10.1080/10255842.2019.1584795.

Ghaneei, Maliheh; Komeili, Amin; Li, Yong; Parent, Eric C.; Adeeb, Samer (2018): 3D Markerless asymmetry analysis in the management of adolescent idiopathic scoliosis. In: BMC musculoskeletal disorders 19 (1), S. 385. DOI: 10.1186/s12891-018-2303-4.

Komeili, Amin; Westover, Lindsey; Parent, Eric C.; El-Rich, Marwan; Adeeb, Samer (2015): Correlation Between a Novel Surface Topography Asymmetry Analysis and Radiographic Data in Scoliosis. In: Spine deformity 3 (4), S. 303–311. DOI: 10.1016/j.jspd.2015.02.002.

Lausser, Ludwig; Siegle, Lea; Rothbauer, Wolfgang; Frank, Derk; Just, Steffen; Kestler, Hans A. (2018): Semantic Multi-Classifer Systems Identify Predictive Processes in Heart Failure Models across Species. In: Biomolecules 8 (4). DOI:10.3390/biom8040158.

Lausser, Ludwig; Szekely, Robin; Schmid, Florian; Maucher, Markus; Kestler, Hans A. (2022): Efficient cross-validation traversals in feature subset selection. In: Scientific reports 12 (1), S. 21485. DOI: 10.1038/s41598-022-25942-4.

Qi, Charles R.; Su, Hao; Mo, Kaichun; Guibas, Leonidas J. (2016): PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation.

Rothstock, Stephan; González-Ruiz, José María; Weiss, Hans-Rudolf; Turnbull, Deborah; Krueger, Daniel (2023): Severity and Cobb angle of scoliosis patients quantified with markerless trunk surface topography using k-NN search and multivariate regression analysis. In: Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, S. 1–7. DOI: 10.1080/21681163.2023.2238840.

Rothstock, Stephan; Weiss, Hans-Rudolf; Krueger, Daniel; Kleban, Victoria; Paul, Lothar (2020): Innovative decision support for scoliosis brace therapy based on statistical modelling of markerless 3D trunk surface data. In: Computer methods in biomechanics and biomedical engineering 23 (13), S. 923–933. DOI: 10.1080/10255842.2020.1773449.

Kontaktdaten

AI4EA

eMail: AI4EA@gfai.de

Internet: www.gfai.de/aktuelles/veranstaltungen

3D-NordOst:

eMail: 3d-nordost@gfai.de

Internet: 3d-nordost.de

IMPRESSUM

Herausgeber / Organisation:

St. Rothstock, B. Hohnhäuser, M. Koddenbrock, B. Botsch, M. Bauer

Gesellschaft zur Förderung angewandter Informatik e. V. (GFai)

Volmerstraße 3, 12489 Berlin

Telefon: +49 30 814563-300

Telefax: +49 30 814563-302

eMail: info@gfai.de

Internet: www.gfai.de

ISBN-Nummer:

978-3-942709-32-3