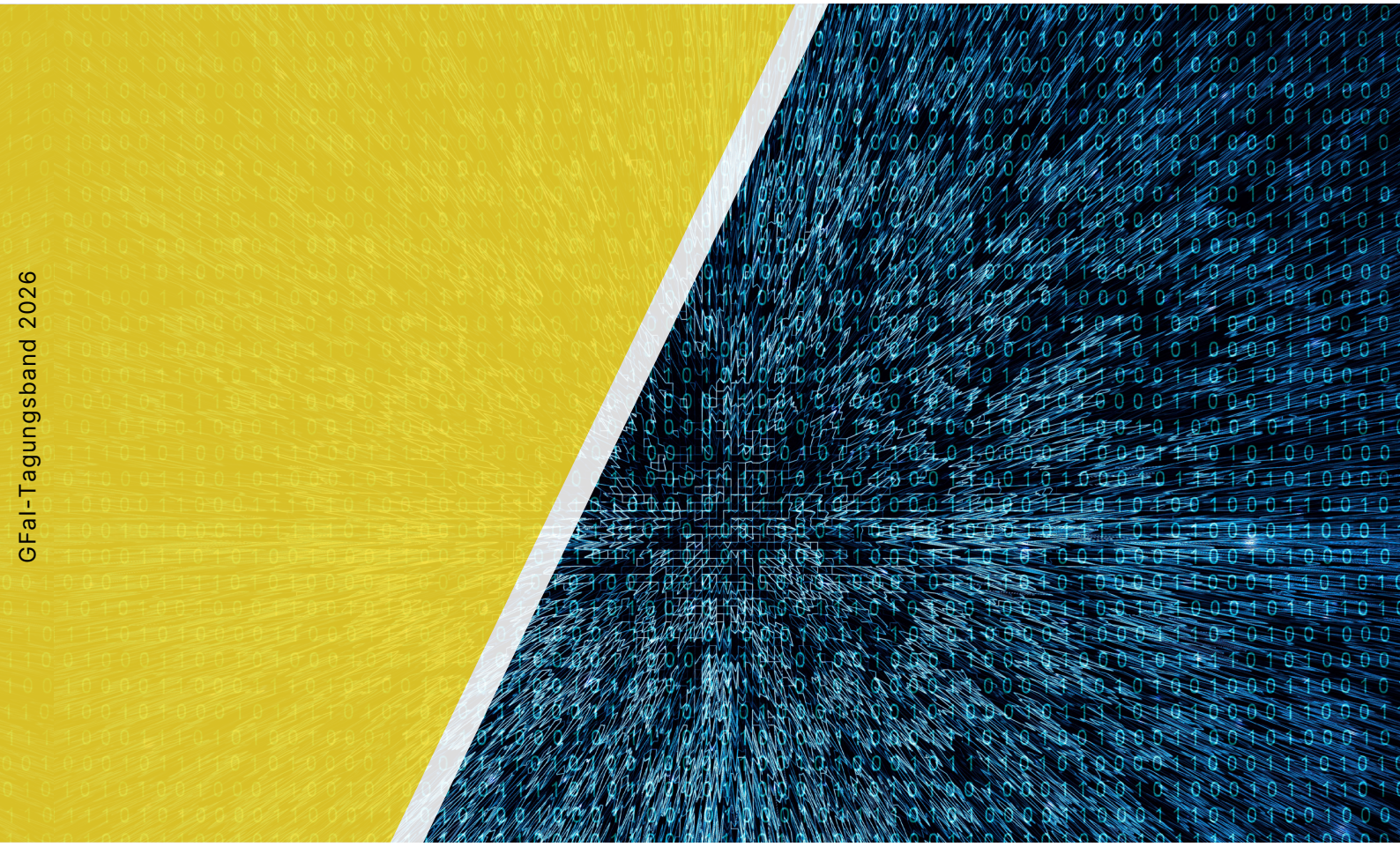
GFaI Future Tech Days

Tagungsband 2026



GFaI-Tagungsband 2026

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

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

Gesellschaft zur Förderung
angewandter Informatik e.V.

GFaI Future Tech Days

Tagungsband 2026

INHALTSVERZEICHNIS

GFaI-Tagungsband 2026

VORWORT

3 Vorwort der Workshopleitung

ASSISTENZ, MONITORING UND SICHERHEIT

4 EnaHuf – an AI powered embedded system for remote monitoring of horses

11 Was ich mir als Autofahrer bei der Bootsnavigation wünsche – Konzeption eines optischen Assistenzsystems für die Sportschifffahrt

17 Aktueller Arbeitsstand zur Patientenpositionierung in der Strahlentherapie - geometrische Systemmodellierung des Demonstrators

KI IN SIMULATION UND FERTIGUNG

23 From Data-Driven to Physics-Informed: Benchmarking Neural Surrogate Architectures for Transient Thermal Prediction in Industrial Casting

30 Additively Manufactured Sheet Metal Tools for High-Strength Steel Processing

3D-ERFASSUNG, REKONSTRUKTION UND KALIBRIERUNG

36 Per-Dot Calibration Approach for Single-Shot 3D Sensors Using Structured Light

43 Markerbasiertes Trackingsystem zur roboterunabhängigen Trajektoriengenerierung für automatisierte Beschichtungsprozesse

49 Entwicklung eines Rückenmodell-Generators zur robusteren Klassifizierung von Skoliose

56 Photogrammetric 3D Model-Based Animal Behavior Classification

3D-INSPEKTION, DEFEKT UND REPARATUR

64 Automated robust 3D defect reconstruction for additive repair processes

70 Evaluation of an AI-Based Concept for 2D-to-3D Defect Detection in Industrial Selective Repair Processes

76 Methoden zur Qualitätssicherung urbaner digitaler Zwillinge

MACHINE LEARNING UND DATENGETRIEBENE METHODEN

82 Detection of Anomalous Energy Consumption Patterns in Industrial Machines through Time Series Analysis to Improve Energy Efficiency

89 From LSTM Baselines to Edge-Deployed Mixture-of-Experts: Five Years of Neural Automatic Passenger Counting

95 RAISSE-Methodology for Interpretable Transfer Learning in Restricted Environments with an Illustrative Example on Gait Phase Classification

101 A data-driven optimization approach towards better automation in industrial fused granular fabrication

ZUKUNFTSTECHNOLOGIEN UND KI-SYSTEME

107 Dark Company: Autonomous Multi-Agent Systems and the Decoupling of Corporate Output from Headcount

113 OpenUSD as a Semantic Framework for GMP-Compliant Automated Optical Inspection

VORWORT

Liebe Kolleginnen und Kollegen,
wir freuen uns, Ihnen den Tagungsband der GFaI Future Tech Days 2026 vorzulegen:
3D-iSA & AI4EA: 3D-Datenverarbeitung und KI im industriellen Einsatz.

Am 11. und 12. November 2026 finden die GFaI Future Tech Days zum zweiten Mal in Berlin-Adlershof statt. Im vergangenen Jahr haben wir die Workshopreihen 3D-in-Science & Applications (3D-iSA) und AI for Engineering Applications (AI4EA) erstmals unter einem gemeinsamen Dach zusammengeführt. Mit der zweiten Ausgabe wird aus diesem ersten gemeinsamen Schritt ein festes Format.

Ausgangspunkt der Zusammenführung war die Beobachtung, dass die Grenzen zwischen 3D-Datenverarbeitung und künstlicher Intelligenz zunehmend verschwimmen. Die Beiträge in diesem Band spiegeln diese Entwicklung wider: Viele Arbeiten lassen sich nicht mehr eindeutig einem der beiden Bereiche zuordnen, sondern verbinden die Erfassung und Analyse dreidimensionaler Daten mit lernenden Verfahren.

Alle in diesem Band veröffentlichten Beiträge wurden von Mitgliedern des Programmkomitees in einem zweistufigen Begutachtungsverfahren geprüft, um wissenschaftliche Qualität und Relevanz sicherzustellen. Das Programm gliedert sich in sechs thematische Sessions. Der erste Tag setzt seinen Schwerpunkt auf die 3D-Datenverarbeitung, der zweite auf die künstliche Intelligenz, wobei die Übergänge zwischen beiden Tagen fließend sind.

Ergänzt wird das Programm durch Pitches von Studierenden, die Einblicke in aktuelle Arbeiten des wissenschaftlichen Nachwuchses geben.

Unser besonderer Dank gilt allen Autorinnen und Autoren für ihre Einreichungen sowie dem Programmkomitee für die Begutachtung und die inhaltliche Gestaltung des Programms. Ebenso danken wir dem Organisationsteam für die Vorbereitung der Veranstaltung, den Studierenden für ihre Pitches und allen Teilnehmenden, die mit ihren Ideen, Fragen und Diskussionen zum Gelingen der Future Tech Days beitragen. Für ihre wertvolle Unterstützung danken wir außerdem Frau Beuster vom GRW-Netzwerk NET4AI.

Wir wünschen Ihnen eine anregende Lektüre und freuen uns auf den Austausch während der Veranstaltung.

Berlin-Adlershof, Oktober 2026 Felicitas Böhm



Programmkomitee

Prof. Alfred Iwainsky (GFaI e.V.)
Benjamin Hohnhäuser (GFaI e.V.)
Benny Botsch (GFaI e.V.)
Dr. Christof Puhle (GFaI e.V.)
Daniel Krueger (GFaI e.V.)
Prof. Erik Rodner (HTW Berlin)
Felicitas Böhm (GFaI e.V.)
Prof. Frank Neumann (HTW Berlin)
Habib Ben Khedher (GFaI e.V.)
Prof. Holger Schlingloff (Fraunhofer FOKUS and Humboldt University)
Prof. Jörg Reiff-Stephan (TH Wildau)
Dr. Jürgen Jakumeit (Access e.V.)
Mario Koddenbrock (HTW Berlin)
Martin Wolff (GFaI e.V.)
Leonardo Maben (GFaI e.V.)
Orhan Stephan (GFaI e.V.)
René Heidenreich (GFaI e.V.)
Shaunak Kanikar (GFaI e.V.)
Steven Behm (GFaI e.V.)
Prof. Wolfgang Halang (Fernuniversität in Hagen)

EnaHuf: an AI-powered embedded system for remote monitoring of horses

Erik Haarländer[1], Michael Tebbe[1], Nicolas Grube[1], Anna-Luise Böhm[2], Jan Sprung[3]

[1]ITPower Solutions GmbH, Kolonnenstr. 26, 10829 Berlin,
{Erik.Haarlaender, Michael.Tebbe, Nicolas.Grube}@itpower.de, www.itpower.de

[2]Frankenförder Forschungs GmbH, Potsdamer Str. 18A, 14943 Luckenwalde,
boehm@frankenfoerder-fg.de, www.frankenfoerder-fg.de

[3]ZeSys e. V. Zentrum zur Förderung eingebetteter Systeme, Wagner-Régeny-Straße 16, 12489 Berlin,
sprung@zesys.de, www.zesys.de

Abstract: This paper presents EnaHuf, an AI-powered embedded system integrated into a polymer composite horseshoe for remote horse monitoring and theft prevention. To achieve energy autonomy, the system utilizes a membrane-type electromagnetic energy harvester (4 mWs per step), extending the lifespan of a 150 mAh coin cell to approximately six weeks. Behavioral classification, specifically the detection of guided movement and trailer loading, is performed on-device using a Temporal Convolutional Network (TCN) on a Nordic nRF54L15 microcontroller. The Float32 baseline achieved 71.1 % accuracy on synthesized theft sequences, dynamic range quantization drastically reduced the memory and compute footprint while only reducing accuracy to 69.7 %, a 2 % relative decrease. However, our target metric recall for the critical "Theft (trailer)" class stayed consistent with 83.3 % for both models. These results validate the viability of hoof-integrated theft monitoring.

Keywords: AI; Animal; Energy Harvesting; Embedded Systems; Horse

1 Introduction

This paper reports the latest developments of "EnaHuf - Self-powered sensor-based system for remote monitoring of ungulates" a project from the AI4Tech network[1] , focussing on key aspects during development of the smart horseshoe prototype. In Germany, approximately 1.25 million horses are kept, mostly on expansive, unsupervised pastures. Due to their domestication, these animals are accustomed to humans, significantly increasing the risk of theft. Given the high economic and personal value of these animals, the loss of a horse places a considerable emotional and financial strain on their owners. Therefore, there is a critical need for a reliable system capable of continuous monitoring, early detection of unusual behavior, and immediate emergency alerting in unsupervised environments.

The horse's hoof provides an ideal location for such a system. Horses commonly wear shoes for joint protection on hard ground, making the horseshoe a natural, discreet integration point that cannot be immediately identified as a tracking device by unauthorized persons and cannot be removed quickly without the need for specialist tools. Furthermore, hoof movement is intricately tied to the animal's overall behavior, providing a high-density data source for activity classification. However, implementing a tracking system within a horseshoe presents several unique engineering challenges: (1) Durability is paramount; the electronics must withstand extreme mechanical stresses and harsh environmental conditions, including moisture, mud, and large temperature ranges. (2) Energy autonomy is critical, as horseshoes are typically replaced every 6 weeks. (3) Reliable connectivity is required to transmit data from remote pastures to owners worldwide, necessitating a robust wireless antenna solution and a scalable infrastructure. (4) The system must be intelligent and user-centric, distinguishing between normal movement and suspicious activity (e.g., unauthorized transport) while remaining seamless and easy for the owner to operate.

The EnaHuf project addresses these requirements through a multi-component prototype. The core is a "smart horseshoe" (the EnaHuf), a polymer shoe housing an integrated gyroscope, accelerometer, and Bluetooth Low Energy (BLE)

[1]https://www.ai4tech.de/

chip, supplemented by energy-harvesting hardware. This device runs an embedded deep learning model to classify behavior locally at the edge.

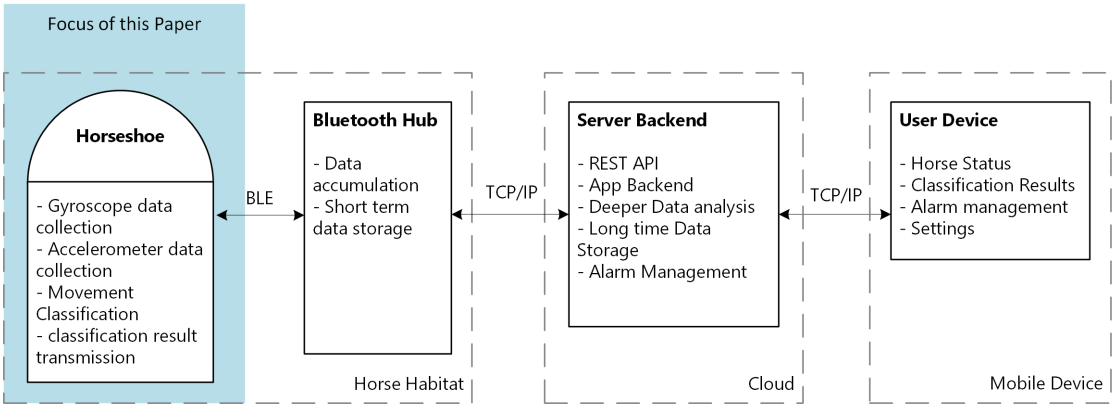


Fig. 1: Architecture of the distributed alarm system showing the components, their intended functions, and their interfaces.

To enable worldwide monitoring, the horseshoe is integrated into a distributed architecture (Fig. 1). Data is transmitted via BLE to local hubs, installed along the horse habitat and relay the information via TCP/IP to a cloud-based backend for storage, analysis, and alarm management before being pushed to the user's device. This tiered approach minimizes power consumption at the sensor level while maintaining scalability. While the EnaHuf ecosystem is comprehensive, this paper focuses specifically on the development and validation of the EnaHuf itself. We highlight two key contributions:

1. **Hardware Design and Energy Harvesting**: The development of a sensor-integrated polymer horseshoe and a proof-of-concept analysis comparing harvested kinetic energy against the system's power requirements.
2. **Embedded AI**: The design and performance evaluation of a Temporal Convolutional Network (TCN) for behavior classification, including an analysis of the trade-offs introduced by quantization.

2 State-of-the-Art

Livestock tracking and theft detection is a wide field. Different products are available today, often as tracking devices incorporated into a collar. While such devices enable long-term tracking of livestock and other animals by use of large batteries, they are obtrusive and easily catch the eye of a potential thief [1–5]. Instrumented measurement of equine movement is well established, particularly for lameness diagnosis, load analysis, and performance evaluation. In addition to optoelectronic methods, inertial measurement units (IMUs) attached to the hoof are used for this purpose. A comparative study showed that a hoof-mounted IMU can record key temporal movement parameters and stride length during walk and trot with a high degree of agreement with an optoelectronic reference method [6]. Commercially available systems include, e.g., the Werkman BLACK system [7] and the Consteed Smart Horse Sneaker by Continental [8], which record hoof movements at a high sampling rate and analyze, in particular, the contact, stance, roll-off, and swing phases. However, such systems are powered by large batteries; energy autonomy derived from the horse's movement is not a feature of the known solutions.

The measurement of forces and pressure distributions under the hoof is also the subject of extensive research. In particular, resistive pressure sensors are used for this purpose, whose electrical resistance changes under mechanical load. These methods allow for the determination of load distribution but do not provide a power source for the connected electronics. GPS-based systems are well established on the market for tracking horses. These are typically attached to the halter, saddle, or mane and powered by (rechargeable) batteries. This results in both

a visible external attachment and a limited operating time. The solutions under consideration do not describe the integration of tracking and communication electronics into the horseshoe. The use of bio-mechanical kinetic energy to autonomously power electronic systems has been scientifically investigated in humans. Donelan et al. developed a knee-mounted generator that generated an average of approximately 5 W of electrical power while walking by utilizing negative joint work [9, 10]. Further studies confirm the potential of biomechanical energy-harvesting systems but describe their application exclusively in humans [11, 12]. The reviewed literature does not document the application of this principle to the movement of the horse's hoof or its integration into a horseshoe. This results in the research gap addressed by this project: the combination of permanent integration on the horse, theft detection, integrated electronics, and autonomous power supply through micro-energy harvesting from hoof movement.

3 Hardware Design and Energy Harvesting

We based our prototype on a modified polymer composite horseshoe model Duplo by H. Frank Kunststofftechnik GmbH (see fig. 2). The Energy Harvesting proof-of-concept analysis compares the energy requirements of the tracking system against the energy harvested by the prototype. The horseshoe design needs to allow for a shoeing surface area of up to 20 cm^2 with a maximum height of 13 mm, and to withstand static loads of over 400 kg and dynamic loads of over 1,000 kg per shoe, which the electronic hardware needs to be protected from. The mechanical service life is to be at least six weeks, whilst the target mechanical lifespan is more than twelve weeks. The intended operating temperature range is −20 °C to +45 °C; an IP69 protection rating is targeted for the electronics. To accomplish these requirements, the construction concept is based on three functionally coordinated thermoplastic polyurethane (TPU) layers. A first soft layer that touches the hoof with Shore A hardness of approximately 50–70 provides cushioning and mechanical decoupling for the sensor system. A second TPU layer with Shore A hardness of approximately 70–85 provides added protection and load-bearing functions. The outer layer is made of wear-resistant TPU with Shore A hardness of approximately 90–98.

The electronic system consists of a nrf54l15 microcontroller unit (MCU) from Nordic Semiconductors for computational tasks, with an integrated BLE transceiver for wireless data transfer, a Bosch BMI270 Inertial Measurement Unit (IMU) to collect movement data, a rechargeable battery attached to a micro-energy harvesting unit for power supply, and a customised antenna. Data transmission via BLE, MCU and IMU typically require at least around 100 µA each when active. The EnaHuf therefore needs optimised hardware for minimal power consumption and a reliable power supply in a very confined space. To overcome this challenge a micro energy harvesting solution capable of converting kinetic energy into electrical energy is used.

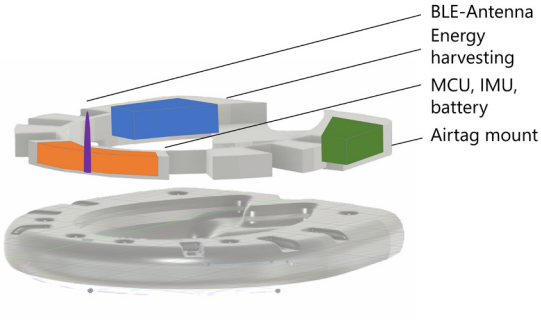
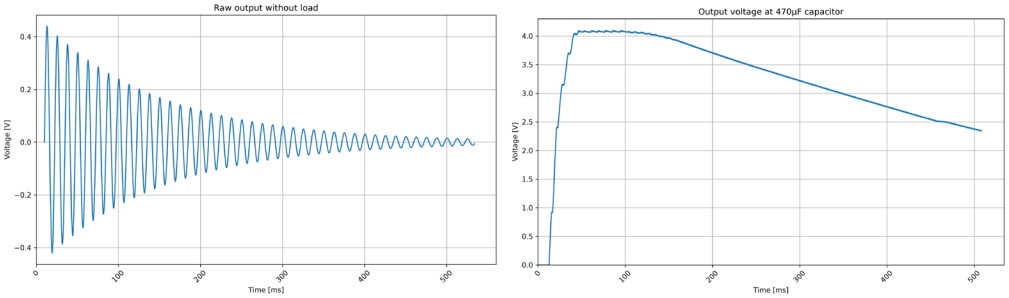


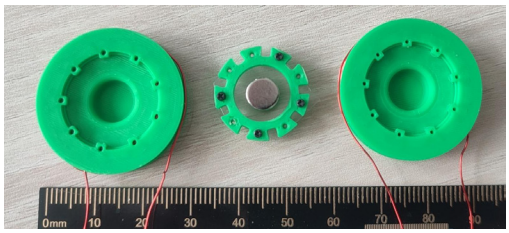
Fig. 2: The coloured parts show how the hardware will be integrated into the horseshoe.

The chosen energy harvesting approach implements a membrane-type electromagnetic harvester, which operates via a magnet affixed to a flexible membrane that oscillates relative to a stationary coil similar to a design found in [13] (see fig. 3 (c)). This configuration enables a comparatively low space requirement, making it ideal for integration into horseshoes. However, the system's primary drawback is its inherently low voltage output, which necessitates careful optimization of parameters such as membrane stiffness and magnetic flux density to achieve practical energy levels. Its main energy input occurs precisely when the hoof impacts the ground, converting the

mechanical shock into vibration of the membrane. For each impact, the energy of about 4 mJ (see fig. 3 (a),(b)) can be stored in a rechargeable coincell. With an assumed mean horse step count of 1 Hz we expect 14.4 J of harvested energy per hour, which would sum up to about 170 J per day if the horse is active 12 hours per day. The energy harvesting connected to a fully charged coincell of about 150 mAh at 3 V (1620 J) increases its energy capacity by roughly 10 % and enables the EnaHuf-horseshoe to run for about six weeks.



(a) (b)



(c)

Fig. 3: (a) shows the raw output without load of one step. (b) the output voltage at a capacitor of 470 µF. The voltage goes up to 4 V resulting in about 4 mWs energy harvested at one step or about 1 mA in average at 4 V. The data was collected on a mechanical simulator. (c) Energy harvesting prototype disassembled.

4 Embedded AI

Dataset Acquisition and Preprocessing To train the classifier, we collected 6-axis IMU data (3D translational acceleration in $\frac{m}{s^2}$ and 3D rotational velocity in $\frac{rad}{s}$) sampled at 500 Hz. The data was gathered by equipping a horse with the physical prototype while simultaneously observing and labeling the animal's behavior. The IMU is configured with automatic offset correction to stabilize acceleration and gyroscope data. Both acceleration and gyroscope data are then used for classification.

We define five distinct classes: *Standing (Free)*, *Standing (Guided)*, *Walking (Free)*, *Walking (Guided)*, and *Theft (trailer)*. We explicitly distinguish between "Free" and "Guided" states to detect the presence of a human, which is often the first indicator of an attempted theft. The "Theft (trailer)" class specifically identifies the process of a horse entering a trailer, the most common method of horse theft. Other theft cases, in which the horse leaves its habitat, are covered by Bluetooth distance measurement and are not described in this paper.

To ensure model robustness against sensor mounting variances and alignment errors, the raw data underwent three-dimensional gravity normalization. The mean gravity vector was calculated over the entire measurement and transformed via a rotation matrix to align the static gravity component with the positive Z-axis (the surface normal of the EnaHuf). Following channel-specific scaling, the time series were split at temporal discontinuities (Δt<0.5 ms or >3.5 ms) and divided into non-overlapping sliding windows of $N = 200$ data points. The resulting dataset of 8,693 windows was split into a balanced test set (250 samples) and a training/validation set (8,443 samples), with a strict 80/20 split for the latter to monitor convergence and to prevent overfitting.

Model Architecture Deploying real-time classification on a microcontroller (Nordic nRF54L15, ARM Cortex-M33, 256 KB RAM) necessitates an architecture that balances predictive power with minimal memory overhead. We evaluated several candidates including Transformer [14–16] and LSTM architectures [17–20], but decided on Temporal Convolutional Networks (TCNs) for their fast inference and a small memory footprint and high performance, even in quantized formats [19, 20]. We employed a TCN utilizing TensorFlow (2.21.0) and Keras (3.13.2). The architecture consists of dilated 1D convolutional layers to capture both short-term dynamics (e.g., hoof strikes) and longer-term periodic gait patterns without the need for recurrent buffers. To stabilize convergence and facilitate subsequent quantization, we utilized Batch Normalization and ReLU6 activations. To further reduce the parameter count and prevent overfitting, we replaced dense hidden layers with a Global Average Pooling (GAP) layer, followed by a Softmax output layer for the five gait states. Table 1 shows our chosen architecture.

Table 1: Our TCN Architecture parameters as described in the *Details* column, L_2 refers to L_2-Regulation

Layer Name	Type	Output Shape	Parameters	Details
input_layer	InputLayer	(200, 6)	0	200 samples $\times$ 6 IMU channels
conv1d + relu	Conv1D, ReLU6	(200, 64)	1984	64 filters, kernel 5, stride 1, $L_2 = 10^{-4}$
conv1d_1 + relu_1	Conv1D, ReLU6	(100, 64)	20544	64 filters, kernel 5, stride 2, $L_2 = 10^{-4}$
conv1d_2 + relu_2	Conv1D, ReLU6	(50, 32)	10272	32 filters, kernel 5, stride 2, $L_2 = 10^{-4}$
conv1d_3 + relu_3	Conv1D, ReLU6	(50, 32)	3104	32 filters, kernel 3, stride 1, $L_2 = 10^{-4}$
global_avg_pool	GlobalAvgPool1D	(32)	0	Average pooling over time dimension
dropout	Dropout	(32)	0	Rate $p = 0.25$
dense	Dense, Softmax	(5)	165	5 classification outputs

Training and Quantization To address the inherent class imbalance of field data, the model is trained using a Categorical Cross-Entropy loss function with inverse class weights. To minimize critical false negatives in theft scenarios, the weight for the "Theft (trailer)" class is further amplified by a factor of 3.0. We employ the Adam optimizer (η=10^{-3}, batch size 32) over 120 epochs, utilizing adaptive callbacks for learning rate reduction and early stopping. During training, the loss value converged to a minimum validation loss of $\mathcal{L}_{val} = 0.5918$, achieving an internal validation accuracy of 80.7% at epoch 53 (see fig. 4).

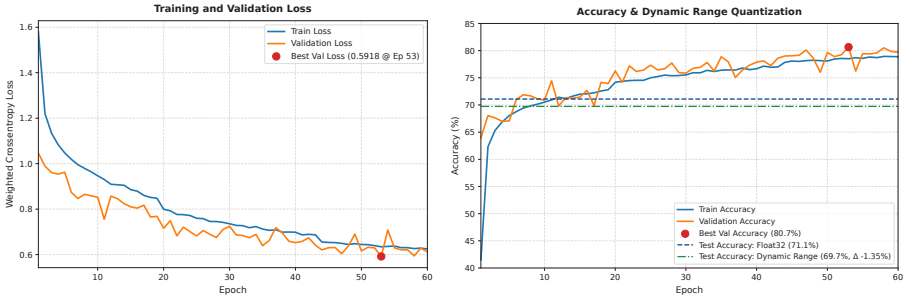


Fig. 4: Training and validation metrics of the base Float32 model over 120 epochs.

To minimize CPU active time during energy harvesting cycles and memory usage, the model is converted to dynamic range quantized TensorFlow Lite Micro model, utilizing FLOAT32 input/ output format and INT8 for internal weights and activations. A primary challenge during this process was the scale mismatch between the accelerometer ($\pm 220\ \frac{\mathrm{m}}{\mathrm{s}^2}$) and the gyroscope ($\pm 8\ \frac{\mathrm{rad}}{\mathrm{s}}$). Standard global quantization causes smaller gyroscope signals to round to zero, severely degrading accuracy. We resolved this by applying channel-specific static pre-scaling factors, normalizing each raw feature channel to the range [−1.0,1.0]. This ensures that the 8-bit integer range [−128,127] is fully utilized across all six channels.

Performance Evaluation: Baseline versus Quantized Model We compared the baseline FLOAT32 model to the dynamic range quantized model in terms of resource usage and scenario-based reliability. Quantization reduced the storage footprint from 148 KB to 49 KB (a 67 % reduction). Furthermore, the shift to integer arithmetic significantly

reduced inference latency on the Cortex-M33 core, enabling 80 millisecond inference time.
To evaluate the model under continuous, realistic conditions, we developed a custom validation environment that concatenates raw data segments into a "stealthy theft sequence": *Free Standing/Walking* → *Guided Standing* (horse approached by thief) → *Guided Walking* (lead away) → *Theft (trailer)* (trailer). Gaussian noise was injected to generate variance in data. Across 100 iterations, the average overall accuracy decreased from 71.1 % (FLOAT32) to 69.7 % (quantized), which corresponds to an absolute loss of 1.35 percentage points (a 2.0 % relative degradation). Fig. 5 compares the class-wise confusion matrices of the unquantized and quantized models.

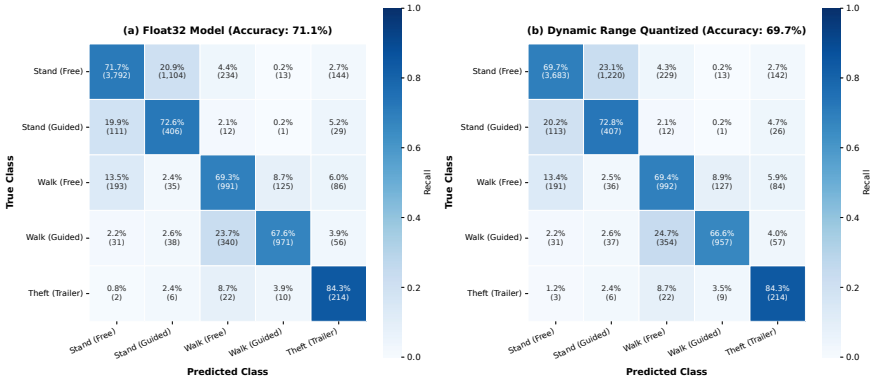


Fig. 5: Confusion matrices comparing the FLOAT32 baseline model (left) and the dynamic range quantized model (right) averaged over 100 simulated theft sequences.

5 Discussion

The AI model is quantized to enable fast latency and minimal memory footprint on embedded hardware. However, this leads to a performance penalty. In the context of theft prevention, the "Theft (trailer)" class is most critical. The false negative rate (FNR) should be minimal, to ensure robust theft detection. Fig. 5 shows, that our model exhibit a false negative rate of ∼15.7 % for the "Theft (trailer)" class. We accept a higher FNR in our model, to ensure good quantized performance and implement a second channel for theft detection by Bluetooth distance measurement from horse to local hub to catch false negatives. Since precision for the "Theft (trailer)" class is at 40.9 % we mitigate false alarms through downstream strategies e.g., a classification buffer in the backend's alarm management system. Fig. 5 also shows confusion in free vs. guided situations, which might be due to the familiarity of humans to horses and therefore not be a valid marker of unwanted human approaches. Furthermore the data was collected from one horse in a controlled, supervised environment and only at day time. Field tests with different horses in less controlled and unsupervised environments are necessary to evaluate how well the AI generalizes and to collect yet uncovered horse behaviour e.g., sleeping.
The EnaHuf project set out to provide a solution to the problem of ungulate theft detection for free roaming horses on their pastures. The results show that a theft detection device, integrated into the horse shoe is a viable option to detect theft attempts early. However, the prototype is yet not validated in a field test as a fully integrated system, which is required to measure real-life mechanical resistance, fit the alarm management parameters, evaluate real-world AI classification performance and get representative data on the system's battery life with energy harvesting.

Acknowledgements We are very grateful for our collaboration with H. Frank Kunststofftechnik GmbH and their steady supply of experimental materials. The EnaHuf Project is funded by the German "Wirtschaftsministerium" within the funding program "Zentrale Innovationsprogramm Mittelstand (ZIM)" and originated from the AI4Tech Network. Additionally, we would like to thank Roman Meier and Felix Rohrbach for their contributions.

References

[1] PAJ GPS, "AG GPS tracker für pferde." https://www.paj-gps.de/gps-tracker-fuer-pferde/. Accessed: 2026-07-30.

[2] InfoStars G.m.b.H, "GPS viehortung & GPS sender." https://www.gps-infostars.com/services-produkte/gps-viehortung/. Accessed: 2026-07-30.

[3] UNSER LAGERHAUS, "GPS-ortung für nutztiere mit qtrack." https://lagerhaus.at/unser-lagerhaus/regional-sortiment/a/gps-ortung-fuer-tiere. Accessed: 2026-07-30.

[4] A. Tittel, "Finde deinen GPS-tracker für tiere – ortung & sicherheit jetzt sichern!." https://www.tipronet.net/gps-tracker/weitere-anwendungsgebiete/tiere. Accessed: 2026-07-30.

[5] N. Huang, "Warum LoRaWAN für die GPS-viehverfolgung wählen?." https://www.mokosmart.com/de/why-choose-lorawan-for-gps-livestock-tracking/. Accessed: 2026-07-30.

[6] J. Hagen, F. T. Jung, J. Brouwer, and R. Bos, "Detection of equine hoof motion by using a hoof-mounted inertial measurement unit sensor in comparison to examinations with an optoelectronic technique - a pilot study," vol. 101, p. 103454.

[7] N. Nachrichten, "Sensoren an pferdehufen bieten einen schatz neuer informationen." https://niederlandenachrichten.de/fachartikel/sensoren-an-pferdehufen-bieten-einen-schatz-neuer-informationen/. Accessed: 2026-07-30.

[8] J. Keller, *Validierung und klinische Anwendung eines mit Druck- und Lagesensoren ausgestatteten Hufschuhs bei Pferden*. Cuvillier Verlag, 1. auflage ed.

[9] Continental, "Consteed the smart horse sneaker." https://www.consteed.com/de-de/page-viewer/forriders. Accessed: 2026-07-30.

[10] J. M. Donelan, Q. Li, V. Naing, J. A. Hoffer, D. J. Weber, and A. D. Kuo, "Biomechanical energy harvesting: Generating electricity during walking with minimal user effort," vol. 319, no. 5864, pp. 807–810.

[11] *Ambient energy für ambient intelligence road map des NEMO-Netzwerkes MESEDA zu energy harvesting*. Ges. zur Förderung Angewandter Informatik. OCLC: 310132677.

[12] SolePower, "SolePower." https://solepowertech.com/press. Accessed: 2026-07-30.

[13] F. Khan, F. Sassani, and B. Stoeber, "Nonlinear behaviour of membrane type electromagnetic energy harvester under harmonic and random vibrations," vol. 20, no. 7, pp. 1323–1335.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need." http://arxiv.org/abs/1706.03762.

[15] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training." https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.

[16] J. Yang, J. Liao, F. Lei, M. Liu, J. Chen, L. Long, H. Wan, B. Yu, and W. Zhao, "TinyFormer: Efficient transformer design and deployment on tiny devices." https://arxiv.org/abs/2311.01759v2.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," vol. 9, no. 8, pp. 1735–1780.

[18] A. Ihichr, S. Hakkal, O. Oustous, Y. E. B. E. Idrissi, and A. A. Lahcen, "Extended LSTM to enhance learner performance prediction," vol. 19, no. 4, p. 251.

[19] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling." http://arxiv.org/abs/1803.01271.

[20] B. Saha and R. Samanta, "Rethinking temporal models for TinyML: LSTM versus 1d-CNN in resource-constrained devices." https://arxiv.org/abs/2603.04860. Version Number: 1.

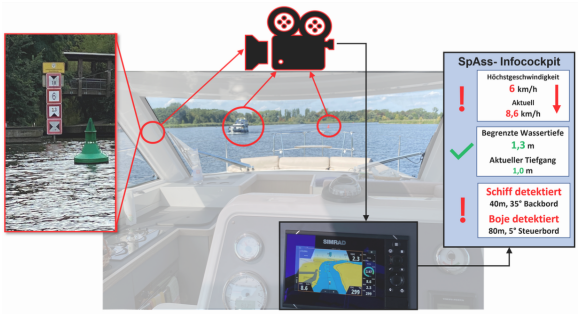
Was ich mir als Autofahrer bei der Bootsnavigation wünsche – Konzeption eines optischen Assistenzsystems für die Sportschifffahrt

Konrad Grützmann, Mirko Riedel
Deveritec GmbH, Bamberger Straße 1, 01187 Dresden, konrad.gruetzmann@deveritec.com,
www.deveritec.com

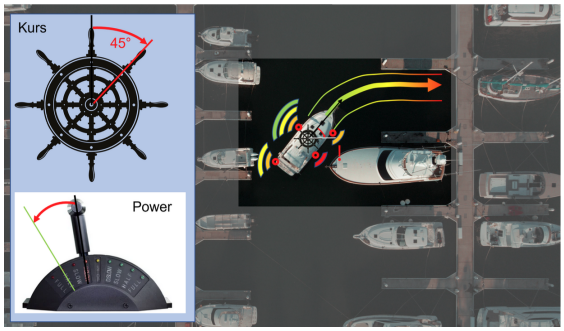
Abstract: In der Sportschifffahrt führen schwierige Navigationsbedingungen und mangelndes nautisches Wissen häufig zu Kollisionen und Sachschäden. Für mehr Sicherheit entwickeln wir daher im Förderprojekt „SpAss – Sportschifffahrtsassistent" ein Assistenzsystem mit kostengünstigem Multi-Kamerasystem, GPS und intelligenter Datenverarbeitung. Wir stellen hier die optische Umfeldsensorik und deren Integration ins Gesamtsystem vor. Zentral sind Objekterkennung und Tracking im Fernfeld (Boote, Schilder) und Hindernisdetektion im Nahfeld für sichere Navigation in enger Umgebung. Zunächst stellen wir die präzise Kamera- und Systemkalibrierung als Systembasis vor. Dann werden beim Panorama-Stitching Bilder verschiedener Kameraperspektiven zusammengefügt. Im Anschluss erfolgen im Panoramabild die Objekterkennung und -segmentierung mittels YOLO26. Schließlich testen wir simultane Kartierung und Lokalisierung (ORB-SLAM3) für Navigation in GPS-schwacher Umgebung. Im Videomaterial einer ersten Testfahrt werden Boote und Personen zuverlässig erkannt und eine rein bildbasiert plausible Bootstrajektorie berechnet.
Keywords: Sportschifffahrt; Navigationsassistenz; Systemkalibrierung; Kamerakalibrierung; Datenfusion; Objekterkennung; Instanzsegmentierung; visuell-inertiales SLAM

1 Hintergrund

In der Sportschifffahrt gibt es viele Verkehrsteilnehmer mit wenig Erfahrung, die sich motorisierte Boote mit teils über 6 m Länge mieten können. Die Steuerung ist schwierig - die Bootsreaktion ist stark zeitlich verzögert zum Steuersignal, Größen und Abstände sind schwer einschätzbar. Verkehrsschilder und -regeln sind dem üblichen Autofahrer wenig vertraut. Kurzum, Orientierung und Navigation sind sehr ungewohnt und führen mitunter zu versicherungsrelevanten Schäden. Unser Ziel ist es, ein System zu entwickeln, das den wenig erfahrenen „Kapitän" bei der Navigation unterstützt. Dazu wird ein Bootsmodell mit optischer Umfeldsensorik und -verarbeitung aufgebaut. Dieses erlaubt eine verbesserte Navigation bis hin zu Steuervorschlägen und umfasst auch kritische Elemente wie Brücken, Schleusen und enge Anlegestellen.



Fernfeld: Beispielhafte Erkennung von anderen Schiffen, Bojen, Verkehrsschildern und deren Darstellung für den Bootsführer.



Nahfeld: Assistenz im Nahfeld, z.B. bei kritischen Situationen wie Einparken. Draufsicht mit Steuervorschlägen.

Abb. 1. Beispielhafte Fern- und Nahfelddarstellung in einem Boot einschließlich Informationen und Markierungen für den Fahrer

Wir stellen hier die optische Umfeldsensorik und deren Integration ins Gesamtsystem vor. Objekte sollen verfolgt, sowie Größen und Entfernungen abgeschätzt werden, wobei Nah- und Fernbereich unterschiedliche Anforderungen stellen: 1. Umgebungserkennung im Fernfeld (> 10 m, Schilder, Brücken, andere Schiffe, Abb. 1 links), 2. Hindernisdetektion im Nahfeld (< 10 m, Stege, Bojen, Poller, …, Abb. 1 rechts). Damit wird für die Schifffahrt ermöglicht, was beim Auto Standard ist – die Erkennung von Verkehrszeichen und eine Einparkunterstützung.

2 Aufbau und Einordnung der Umfeldsensorik ins Gesamtsystem

Die Sensorik des Gesamtsystems umfasst Komponenten zur Bootslokalisierung (GNSS, IMU) und die optische Umfeldsensorik (Abb. 2). Letztere besteht aus einem Mehrkamerasystem mit 180°-360° Umblick und ist die

Grundlage zur Erkennung und Verortung relevanter Objekte. Die Position anderer Schiffe wird ergänzend durch das "Automatische Identifikationssystem" erfasst. Basierend auf GNSS entwickelt unser DLR-Projektpartner ein Verfahren, das GNSS-Korrektursignale des High Accuracy Service der Galileo-Satelliten und des SAPOS-Vermessungsdienstes integriert, um eine bis zu zentimetergenaue Eigenposition zu ermöglichen. Ergänzend liefern eine inertiale Messeinheit (IMU) hochaufgelöste Dreh- und Beschleunigungsdaten und, optional, ein Sonar die Bodendistanz. Alle Sensordaten sind mit Zeitstempel, Position und Ausrichtung versehen. Ein zentrales Ziel ist die Entwicklung einer Verarbeitungseinheit, die die Sensordaten algorithmisch synchronisiert und fusioniert sowie die Sensoren durch Transformation relativ zum Boot und zu den Kartendaten verortet (Abb. 2 Mitte). Der Partner TU-Dresden entwickelt Modelle zur Bootsdynamik und ein Modul zur Trajektorienplanung (z.B. für Einparkhilfe), welches die integrierten Daten verarbeitet. Schließlich wird ein Interface entwickelt, das Kurs, Schub, Trajektorie, Schilder, Hindernisse (Boote, Verkehrsschilder, etc.) und eventuelle Warnhinweise visualisiert (Abb. 2 rechts und Abb. 1 beispielhaft).

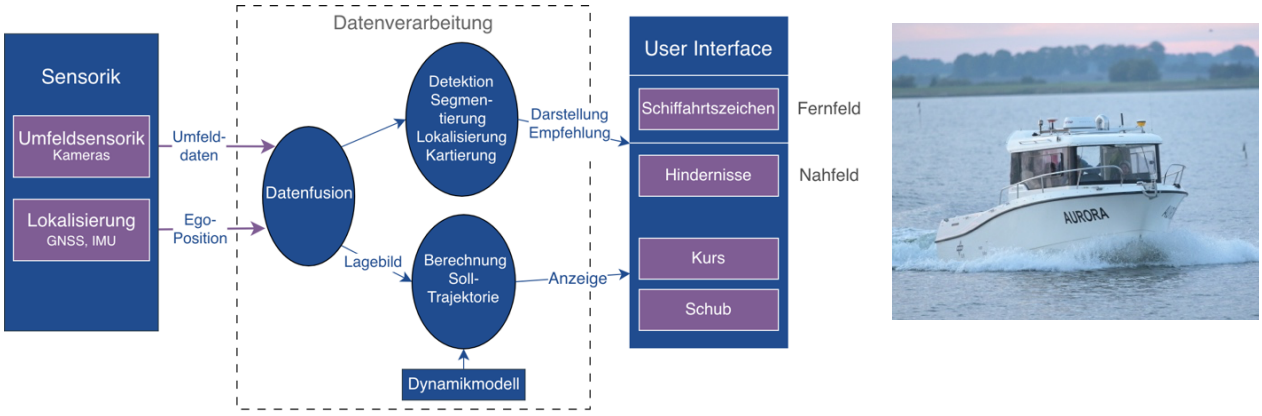


Abb. 2. Links: schematische Darstellung des Gesamtsystems. Rechts: Boot vom DLR-Partner für Messfahrten.

In dem Beitrag wird die Entwicklung der optischen Umfeldsensorik vorgestellt. Deren Hardware soll ein kompaktes, leichtes und preisgünstiges (< 1000 €) System mit Batteriebetrieb sein. Dafür evaluieren wir Systemkategorien mit unterschiedlicher Preisleistung: von Raspberry Pi 5 (3-4 Kameras) bis Industrial PC (IPC, bis 8 Kameras). Alle Varianten nutzen eine eigene, einfache IMU und ggf. einen einfachen GNSS-Empfänger zur Erfassung der systeminternen Relativbewegungen (z.B. Wellengang) und der globalen Position. Wir haben die erste Implementierung der IPC-Kategorie mit drei sehr nah beieinanderstehenden 4K USB-Kameras (Sony IMX415) mit bekannter und fester relativer Orientierung realisiert. Die auf dem Bootsdach nach vorne ausgerichteten Kameras bieten einen Gesamtöffnungswinkel von ca. 140°. Der IPC wird später zwei oder drei dieser Kamerablöcke haben. Aktuell werden mögliche Einschränkungen durch den Rolling-Shutter-Effekt untersucht. Ein zukünftiger Umstieg auf preisintensivere Global-Shutter wäre denkbar.

3 Optische Umfeldsensorik - Methodik und Ergebnisse

Schritte zur Inbetriebnahme sind Kamerakalibrierung (innere Orientierung) sowie Systemkalibrierung (äußere Orientierung aller Sensoren relativ zum Umfeld- und Bootskoordinatensystem). Mit den Systemfunktionen:

- Image-Stitching zum Panoramabild (Fernfeld) und zur Ebene für Draufsicht (Nahfeld)
- Objekterkennung schifffahrtspezifischer Objektklassen, Segmentierung und Tracking
- Objekttracking im Fernfeld: andere Schiffe, Verkehrszeichen etc.
- Simultane Kartierung und Positionierung im Nahfeld: Feinpositionierung zum Manövrieren, Redundanz bei GNSS-Signalproblemen
- Aufbereitung von Panorama und Ebene für Nutzeranzeige

Im Folgenden stellen wir Kamera- und Systemkalibrierung sowie Methoden und Ergebnisse der drei getesteten Funktionen (Panorama-Stitching, Objekterkennung, simultane Positionierung und Kartierung) vor.

3.1 Kamera- und Systemkalibrierung

Kalibrierte Kameras sind essenziell für die Systemkalibrierung und die spätere genaue Ortsbestimmung. Zunächst wurde Software entwickelt, mit der nutzerfreundlich und interaktiv die innere Orientierung (kameraspezifische Brennweite, Hauptpunkt und Verzerrungsparameter, OpenCV Pinhole-Kamera) bestimmt werden kann. Grundlage

für diese Kalibrierung sind Fotoaufnahmen eines bekannten Objekts (Laborraum) mit einfach zu erkennenden Markern (optischer Code) mit genau bekannter Lage. Bei einer Sensorauflösung von 3840×2160 Pixeln kalibrieren wir die Kamera mit einem Rückprojektionsfehler von deutlich unter 1 Pixel. Als Referenzvolumen dient ein exakt eingemessener Raum mit ArUco-Markern mit 3,5×6,0×4,0 m (B×T×H). Die Ortsauflösung der Kameras beträgt bei einem horizontalen Öffnungswinkel (FOV) von ca. 81° in einem Abstand von 10 Metern rund 5mm/Pixel (Sichtfeldbreite $W = 2 \cdot d \cdot \tan\left(\frac{FOV}{2}\right)$ → Genauigkeit $\Delta x = \frac{2 \cdot 10\text{m} \cdot \tan(81°/2)}{3840\ \text{Pixel}} \approx 4{,}48\ \text{mm/Pixel}$.

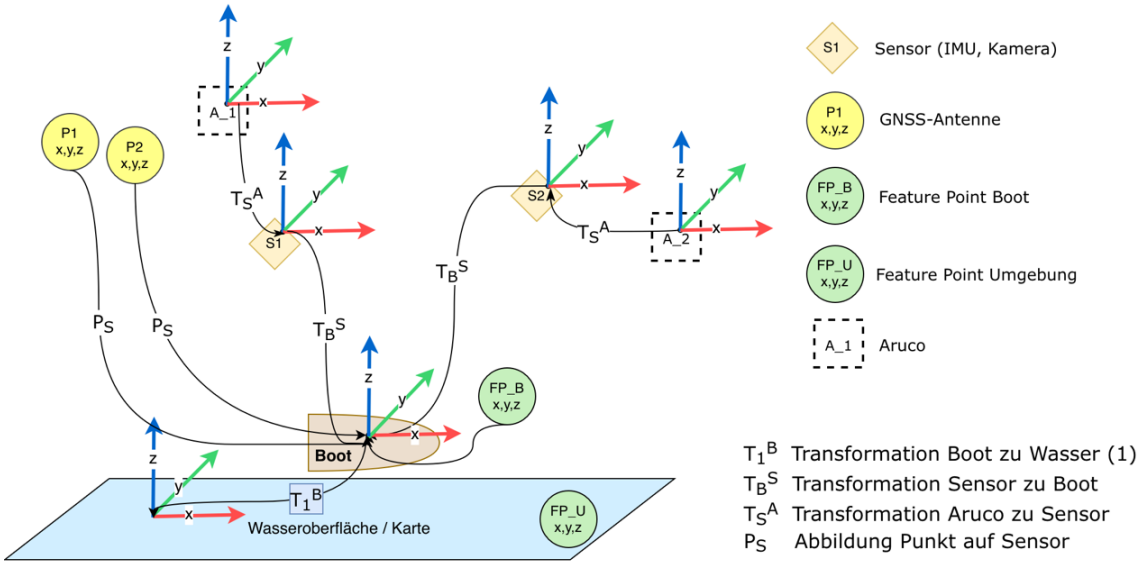


Abb. 3. Schematische Darstellung der Teile des Systems inklusive Sensoren, GNSS-Antennen und Boot. Die jeweilige relative Lage zueinander und zur Wasseroberfläche wird durch Transformationen (Pfeile) definiert.

Bei der Systemkalibrierung werden alle Teile des Systems zueinander mittels Transformationen verortet (Abb. 3). Zur Identifizierung sowie zur Positions- und Orientierungsbestimmung werden an den Sensoren ArUco-Marker so platziert, dass deren relative Lage zum Sensorursprung bekannt ist (Datenblatt). Nach der Installation der Sensoren im Boot wird ein Bilddatensatz aufgenommen und für eine Bündelblockausgleichung auf Basis von Markern und Feature-Points (OpenCV GFTT und ORB) genutzt. Da das Boot auch mit Markern versehen ist und die Wasserlinie sichtbar ist, lassen sich alle Transformationen geschlossen berechnen. Bei einer Objektgröße von 6m wird eine Genauigkeit von ca. ±5mm (FUJI GFX100, ca. 50 Bilder) erreicht.

3.2 Panorama-Stitching

Für Objekterkennung und Fahreranzeige werden die überlappenden Bilder des Multi-Kamerasystems zu einem Panorama (Fernfeld) oder einer Ebene (Nahfeld) fusioniert. Bei nahestehenden Kameras kann näherungsweise ein gemeinsames optisches Zentrum (reine Rotation) angenommen werden. Damit liefert ein einfaches, rein bildbasiertes Panoramastitching mit Schätzung der Rotationshomographie (z.B. OpenCV Stitcher-Klasse) gute Ergebnisse für das Fernfeld (Abb. 4). In der Praxis wird jedes Bild zunächst auf Basis der Kamerakalibrierung entzerrt und dann Feature-Points detektiert (z.B. SuperPoint [1]), die zwischen den Bildern gematcht werden. Zur Erhöhung der Robustheit erfolgt die Ausreißerentfernung durch USAC-Methoden [2,3]. Nun wird die Rotationsmatrix (OpenCV Stitcher) oder die vollständige Kamerapose inklusive Translation mittels eines nichtlinearen Optimierungsverfahrens (z.B. Levenberg-Marquardt) durch Minimierung des Reprojektionsfehlers geschätzt. Die 2D-Bildkoordinaten werden dann unter Verwendung von Pose und innerer Kalibrierung auf einen Zylinder projiziert, der zu einer horizontal unverzerrten 2D-Darstellung "ausgerollt" wird. Zur optischen Verbesserung können Helligkeitsunterschiede der Bilder ausgeglichen und Überlappungsbereiche überblendet werden.
Abbildung 4 stellt ein durch Stitching zusammengefügtes Panorama aus drei Bildern verschiedener Perspektive dar, die auf einer Testfahrt aufgenommen wurden. An den Übergängen der Einzelbilder können noch Artefakte erkannt werden, die in Zukunft durch verbesserte Verfahren behoben werden sollen.
Das beschriebene Live-Stitching ist rechenaufwändig. Wenn die Kameraposen aber fest sind, reicht es, diese einmal zu berechnen und live nur eine feste Transformation anzuwenden (z.B. Lookup-Table für Zielpixel).
Da wir die innere und äußere Orientierung der Kameras durch Vor- und Systemkalibrierung kennen und die Orientierung des Bootes mittels SLAM/IMU (s.u.) live bestimmen können, lässt sich für den Bootsführer eine Vogelperspektive durch Projektion der Kamerabilder auf eine virtuelle Wasseroberfläche generieren (Nahsicht). Die

primäre Herausforderung sind parallaxenbedingte Verzeichnungen von Objekten mit Tiefe (z.B. Poller). Dazu sollen Objektklassifizierung (s.u.) und Tiefenschätzung für eine fehlerarme Bildsynthese integriert werden.



Abb. 4. Beispiel eines Panorama-Stitchings einer Szene der Messfahrt (April 2026, Zierker See, Neustrelitz). Die vereinfachte Rotationshomographie unter Annahme eines gemeinsamen optischen Zentrums (OpenCV *Stitcher*) liefert bereits gute Ergebnisse im Fernfeld. Unten: Ursprungsbilder aus drei Perspektiven. Oben: gestitchtes, randbeschnittenes Panorama.

3.3 Objekterkennung

Ein wichtiger Baustein des Assistenzsystem ist die Objekterkennung – die Lokalisierung aller Objekte auf den Kamerabildern inklusive Bounding-Box, Klassenlabel und Konfidenz. Ein etablierter Algorithmus dazu ist YOLO („You Only Look Once"), der die Erkennung als Regressionsproblem angeht [4]. Das Bild wird in einem Durchlauf durch ein Deep-Learning-Netzwerk geschleust und die Bounding-Boxes etc. mehrerer Objektinstanzen bestimmt. Das neuste Modell der YOLO-Familie ist YOLO26 [5], welches zusätzlich eine Instanzsegmentierung erlaubt. Dabei werden Objektgrenzen pixelgenau erkannt, was für Navigation in enger Umgebung wie Schleusen und beim Einparken unerlässlich ist. Zu Testzwecken haben wir YOLO26 auf dem umfassenden Videomaterial unserer ersten Messfahrt angewendet. Auf den Einzelaufnahmen und auf Panorama-Stitches ist die Erkennung und Segmentierung von Personen und Booten recht zuverlässig (Abb. 5). YOLO26 wurde auf dem MS-COCO-Datensatz [6] mit 80 generellen Objektklassen trainiert wurde. Daher werden viele schifffahrtsrelevante Objektklassen wie Kaimauer, Bojen und Poller nicht erkannt oder fehlklassifiziert (Abb. 5).

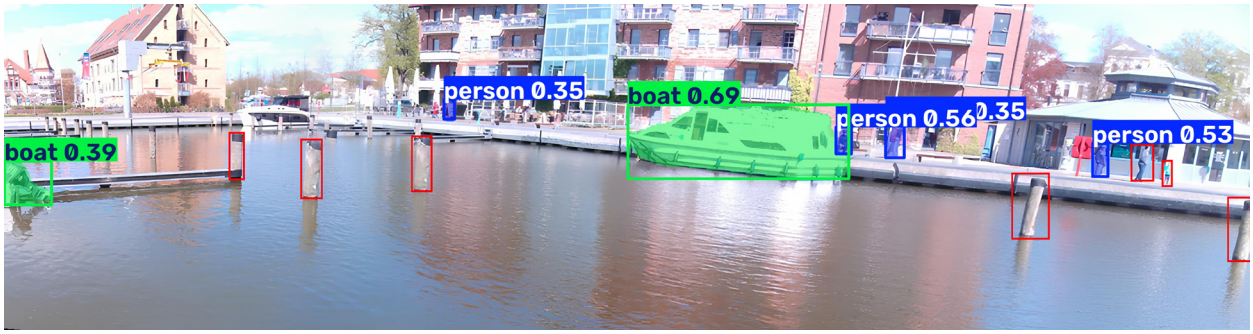


Abb. 5. Objekterkennung mittels YOLO26 liefert Klassennamen mit Konfidenzwerten und Bounding-Boxes sowie pixelgenaue Segmentierung von Booten (grün) und Personen (blau). Grundlage ist oben gezeigtes Panorama, das aus den Bildern der Messfahrt zusammengefügt wurde. Rot: nicht erkannte, manuell annotierte Objekte.

Für den Verlauf des Projekts ist geplant, weitere schiffsrelevante Bilddaten zu sammeln und zu annotieren, mit denen dann die Modelle nachtrainiert werden. Dazu gehören auch Schilder der Binnenschifffahrtsstraßen-Ordnung. Basierend auf den Bilddaten künftiger Fahrten wird z.Z. ein praktikabler Schwellenwert für die Detektionskonfidenz evaluiert (z.B. deutlich größer als 0,35 für Personen in Abb. 5). Dieser soll so gewählt werden, dass das System

hochsensibel relevante Objekte erkennt, ohne eine kritische Menge an Fehlalarmen zu generieren. Durch die zeitliche Konsolidierung der Detektionen lassen sich Recall und Precision steigern. Die Detektionen dienen der informativen Unterstützung, da die finale Entscheidungshoheit beim Bootsführer verbleibt.

3.4 SLAM

Bei Manövern zum Einparken, in Schleusen und unter Brücken kann es zu Abschattung oder Reflektion des GNSS-Signals kommen. Die Methode der simultanen Lokalisierung und Kartierung [7] erlaubt in solchen Situationen eine fortlaufende Navigation allein anhand visueller Merkmale. In den ersten live Umgebungsbildern werden Keypoints bestimmt, die zwischen Bildern konsekutiver Perspektiven gematcht werden. Damit wird eine erste relative Bewegung geschätzt und mit Triangulation eine erste 3D-Karte initialisiert. Fortlaufend werden neue Kamerabilder ausgewertet. Karte und Position werden aktualisiert und z.B. mit Bündelblockausgleich gemeinsam weiter optimiert. Das verringert den Kartenfehler, der sich durch Aufsummieren von Schätz- und Messfehler zwischen den konsekutiven Bildern ergibt. Ein "visuelles Wörterbuch" wird zusammen mit Ähnlichkeitstransformationen von Perspektiven verwendet, um zu prüfen, ob Orte bereits gesichtet wurden. In dem Fall wird die Karte durch Schleifenschluss geschlossen und so der Fehler weiter reduziert. Für erste Tests haben wir ORB-SLAM3 [8] genutzt. Es verwaltet mehrere parallele Karten. Wenn das Tracking verloren geht, wird eine neue Karte generiert. Falls bereits bekannte Bereiche einer anderen Karte auftauchen, werden beide Karten fusioniert. Außerdem erlaubt die visuell-inertiale Variante von ORB-SLAM3 eine sehr schnelle, präzise Initialisierung unter Verwendung von IMU-Daten.

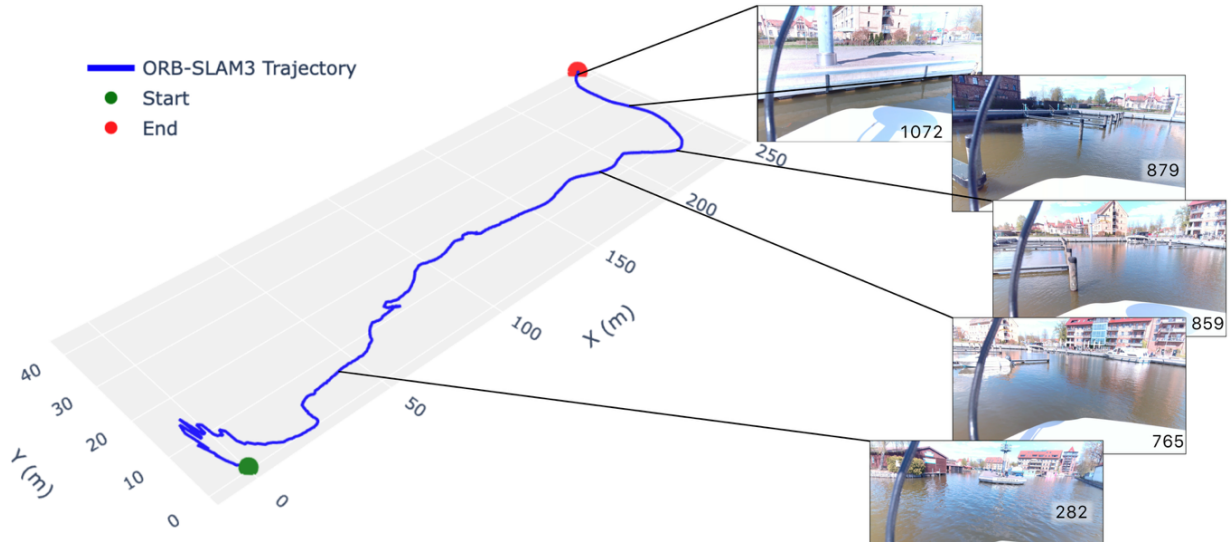


Abb. 6. Links: Trajektorie eines Fahrabschnitts, die mit ORB-SLAM3 aus dem Videomaterial einer Kamera geschätzt und mit Georeferenzen skaliert wurde. Rechts: Auswahl von Bildern inklusive Bildnummer aus diesem Video.

Wir haben die mono-visuelle Variante von ORB-SLAM3 am Videomaterial unserer Messfahrt getestet. Da monokulares SLAM keinen absoluten Maßstab liefert, wurden für die obige Darstellung GPS-Koordinaten von markanten Referenzpunkten im Video bestimmt und die Trajektorie mittels Sim(3)-Alignment (Umeyama-Algorithmus, *evo*-Evaluierungstool [9]) skaliert. Zukünftig wird die korrekte Skalierung direkt online über ein visuell-inertiales Setup (IMU) erfolgen. Abbildung 6 zeigt die Trajektorie eines Teils der Fahrt. Die Strecke passt plausibel zum Videomaterial: am Anfang (grün) fuhr das Boot vom offenen See gerade aus Richtung Hafen (Bildnummer 282, 765). Beim dortigen Anlegen fuhr es eine Linkskurve (859, 879) und schließlich eine kleine Rechtskurve (1072). Anfangs ist die Ungenauigkeit der Schätzung in der Trajektorie sichtbar.

4 Ausblick

Im weiteren Verlauf des Projekts (Ende: Q3/2027) soll das Gesamtsystem hardware- und softwareseitig vervollständigt und getestete Einzelbestandteile umgesetzt werden. Dabei wird die o.g. flexible Systemdefinition umgesetzt und Koordinatentransformationen und Datenfusion implementiert. Die Bilderkennung soll auf dem Panoramabild aus den Kameraperspektiven für den Fernbereich umgesetzt werden. Das Panorama-Stitching soll um die genaue Schätzung der Kameraposen erweitert werden, da sie nicht in jedem Setup exakt das gleiche optische Zentrum haben. Für das Objekttracking soll Bildredundanz und physikalische Kontinuität genutzt werden, um Erkennungsunsicherheiten auszugleichen. Für den Nahbereich werden die Kameraperspektiven zusammengefügt und auf eine Ebene projiziert, die die Draufsicht darstellt. Diese wird zur pixelgenauen

Umfelderfassung durch Bilderkennung verwendet und ist für Navigation in engen Situationen, wie Schleuse oder Einparken, unerlässlich. Einige Kameras haben ein überlappendes Sichtfeld. Eine zusätzlich genaue Abstandsbestimmung durch Triangulation auf Basis dieser Stereobilder soll getestet werden.
Das Nachtraining der Machine-Learning-Modelle soll zunächst an öffentlich verfügbaren Schifffahrtsdaten geschehen, wobei die Modelle um relevante Objektklassen erweitert werden. Bei weiteren Messfahrten wird das System erprobt und praktische Erkenntnisse zur Optimierung genutzt. Es wird weiteres, eigenes Videomaterial gesammelt und annotiert, um die Modelle weiter zu trainieren. Ultralytics' YOLO26 erlaubt ein solches Nachtrainieren der Modelle als auch den Export für spezielle Architekturen. TensorRT wird die Basis für das Deployment auf NVIDIA Orin. ONNX und INT8-Quantisierung wird die Basis für das Deployment auf einem Hailo AI-HAT+ des Raspberry Pi 5.
Weiterhin ist geplant, visuell-inertiales ORB-SLAM3 mit live Kamera- und IMU-Daten zu implementieren. Eine exakte Kamerakalibrierung und die geplante Fusion mit GNSS-Daten erlauben die präzise, kontinuierliche Ortsbestimmung des Boots und bilden die Grundlage für Nahfeldassistenz. Die IMU-Daten ermöglichen genaues Tracking auch an strukturlosen Bildern (unter Brücken, Gegenlicht, starke Kontraste).
SLAM kann auch genutzt werden, um die Sensoren des Systems im Gesamtsystem zu verorten (Systemkalibrierung). Dazu läuft man bei einem neuen Gesamtsetup anfangs einmal um das Boot mit der Kamera. Aruco-Marker auf den Sensoren werden automatisch erkannt und als Keypoints verwendet. Diese Möglichkeit soll getestet werden.

5. Danksagung und Partner

Wir danken dem „Zentralen Innovationsprogramm Mittelstand (ZIM)" für die Förderung unseres Projekts „SpAss – Sportschifffahrtsassistent" im Rahmen des „Innovationsnetzwerks DigiShip – Digitalisierung und Einsatz von Geodaten in der Binnenschifffahrt", Förderkennzeichen: 16KN104302 mit folgenden Projektpartnern:

Deutsches Zentrum für Luft- und Raumfahrt (DLR)
Institut für Kommunikation und Navigation, Abteilung Nautische Systeme, Kalkhorstweg 53, 17235 Neustrelitz, https://www.dlr.de/de/kn/

Technische Universität Dresden
Professur für Informationstechnik für Verkehrssysteme, 01062 Dresden, https://tu-dresden.de/bu/verkehr/vis/itvs

Lange electronic
Lange-Electronic GmbH, Rudolf-Diesel Str. 29, 82216 Gernlinden, www.lange-electronic.com

Referenzen

[1] D. DeTone, T. Malisiewicz, und A. Rabinovich, „SuperPoint: Self-Supervised Interest Point Detection and Description", 19. April 2018, arXiv: arXiv:1712.07629. doi: 10.48550/arXiv.1712.07629.
[2] O. Chum, J. Matas, und J. Kittler, „Locally Optimized RANSAC", in Pattern Recognition, Bd. 2781, B. Michaelis und G. Krell, Hrsg., in Lecture Notes in Computer Science, vol. 2781, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, S. 236–243. doi: 10.1007/978-3-540-45243-0_31.
[3] D. Barath und J. Matas, „Graph-Cut RANSAC", 16. November 2017, arXiv: arXiv:1706.00984. doi: 10.48550/arXiv.1706.00984.
[4] J. Redmon, S. Divvala, R. Girshick, und A. Farhadi, „You Only Look Once: Unified, Real-Time Object Detection", 9. Mai 2016, arXiv: arXiv:1506.02640. doi: 10.48550/arXiv.1506.02640.
[5] G. Jocher, J. Qiu, M. Liu, S. Lyu, F. C. Akyon, und M. E. Kalfaoglu, „Ultralytics YOLO26: Unified Real-Time End-to-End Vision Models", 2. Juni 2026, arXiv: arXiv:2606.03748. doi: 10.48550/arXiv.2606.03748.
[6] T.-Y. Lin u. a., „Microsoft COCO: Common Objects in Context", in Computer Vision – ECCV 2014, Bd. 8693, D. Fleet, T. Pajdla, B. Schiele, und T. Tuytelaars, Hrsg., in Lecture Notes in Computer Science, vol. 8693. , Cham: Springer International Publishing, 2014, S. 740–755. doi: 10.1007/978-3-319-10602-1_48.
[7] R. Mur-Artal, J. M. M. Montiel, und J. D. Tardos, „ORB-SLAM: a Versatile and Accurate Monocular SLAM System", IEEE Trans. Robot., Bd. 31, Nr. 5, S. 1147–1163, Okt. 2015, doi: 10.1109/TRO.2015.2463671.
[8] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, und J. D. Tardós, „ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM", IEEE Trans. Robot., Bd. 37, Nr. 6, S. 1874–1890, Dez. 2021, doi: 10.1109/TRO.2021.3075644.
[9] M. Grupp, „evo: Python package for the evaluation of odometry and SLAM", 2017, [Online], https://github.com/MichaelGrupp/evo

Aktueller Arbeitsstand zur Patientenpositionierung in der Strahlentherapie - geometrische Systemmodellierung des Demonstrators

Riedel, Mirko[1], Konrad Grützmann[1]
[1]deveritec GmbH, Bamberger Str. 1, 01187 Dresden, mirko.riedel@deveritec.com, www.deveritec.com

Abstract: In der Strahlentherapie werden bei Tumoren im Kopfbereich aktuell Fixiermasken eingesetzt, um eine präzise Behandlung des Tumorgewebes zu gewährleisten und das umliegende Gewebe bestmöglich zu schonen. Ein Einsatz eines optischen Trackings an Stelle von harten Fixiermasken wäre aus Patientensicht wichtig, sofern damit das gleiche Maß an Sicherheit gewährleistet werden kann. In Vorarbeiten konnte bereits gezeigt werden, dass dies möglich ist. Aktuell wird ein anwendungsnaher Demonstrator in annähernd Originalgröße aufgebaut, der alle für das Tracking relevanten Komponenten enthält. Kern des Beitrages ist die Vorstellung des Demonstrators mit Fokus auf der Modellierung von Anlagengeometrie und Kameramodell, um das Tracking mit dem höchsten Maß an Sicherheit und Genauigkeit realisieren zu können. Zentrale Themen sind dabei der gewählte Modellansatz sowie die konkrete Modellierung und erste Ergebnisse von Systemkalibrierung und optischen Tracking.

Keywords: Photogrammetrie; geometrisches Systemmodell; Tracking; Systemkalibrierung

1 Einordnung

In der Strahlentherapie werden zur liegenden Bestrahlung von Tumoren im Kopfbereich (Abb. 1, links oben) aktuell Masken zur Fixierung von Patienten (Abb. 1, links unten) verwendet, um durch eine hoch präzise Bestrahlung des Tumors das umliegende Gewebe bestmöglich zu schonen. Die Fixierung ist für den Patienten belastend und kann zu Komplikationen führen. Vor allem beim Übergang zu einer halb aufrechten Behandlung (Abb. 1, rechts) bei der der Patientensitz bewegt wird (Abb. 1, rechts), bietet sich der Wechsel zu einer optischen Überwachung und ggf. der aktiven Nachführung des Patienten an.

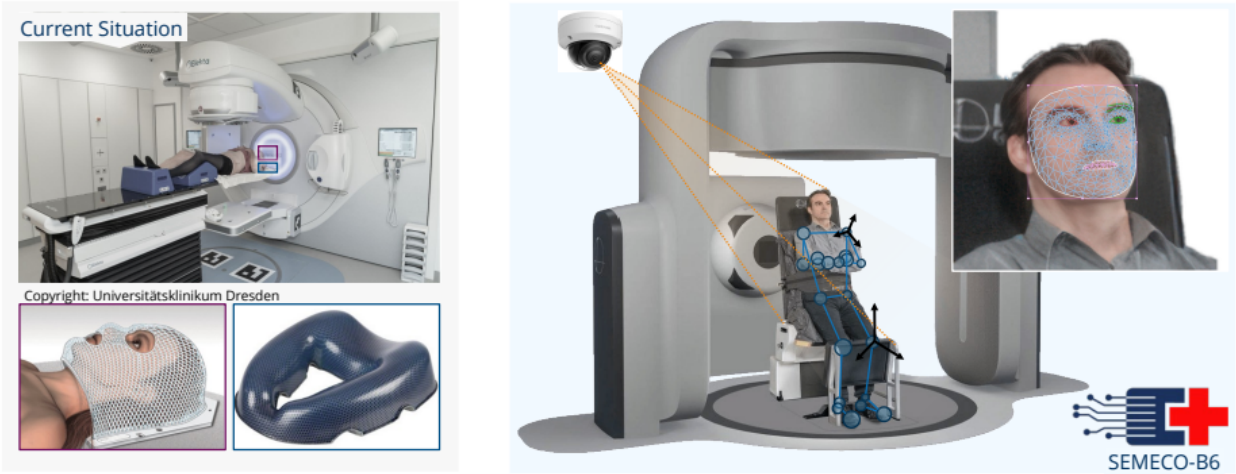


Abb. 1. Links - aktuelle Behandlungssituation, rechts - Übergang zum optischen Tracking

Zum Erreichen der geforderten Genauigkeit beim optischen Tracking sind Kameraanzahl, -anordnung sowie Blickwinkel, Auflösung und Schärfetiefebereich entscheidende Faktoren. Weiterhin muss sichergestellt sein, dass die Sicherheit der Überwachung nicht durch temporäre Verdeckung beeinträchtigt wird. Bei der liegenden Behandlung wird die Strahlquelle um den Patienten rotiert und die Liege geschwenkt. Bei halbaufrechter Behandlung wird der Patient aber relativ zur Strahlenquelle bewegt, so dass Verdeckungen prinzipbedingt recht häufig auftreten.
In vorherigen Publikationen wurden bereits das Systemdesign und unterschiedliche Modelle zur Detektion von Facial-Landmarks vorgestellt und die Tauglichkeit am Funktionsdemonstrator (Abb. 2 links) prinzipiell bestätigt ([1, 2, 3]. Nach den Untersuchungen an diesem verkleinerten, vereinfachten Modell folgt nun die Übertragung auf einen Demonstrator in Originalgröße (Abb. 2 rechts). Dabei ist der aktuelle überwachte Bereich deutlich größer, da jetzt neben der Kopfpose auch die Pose des Tisches miterfasst wird, da dieser während der Behandlung bewegt wird.

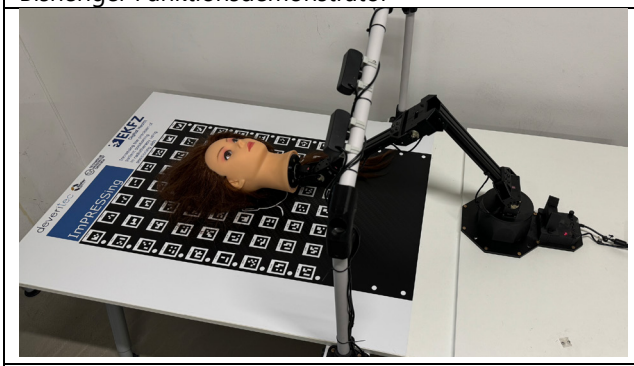
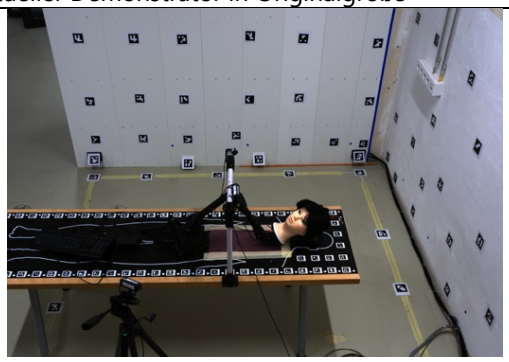
Bisheriger Funktionsdemonstrator	Aktueller Demonstrator in Originalgröße
	
Bewegungsbereich: 500mm x 700mm x 500mm	Bewegungsbereich: 4000mm x 4000mm x 4000mm

Abb. 2. Aufbau und Kenngrößen der Demonstratoren

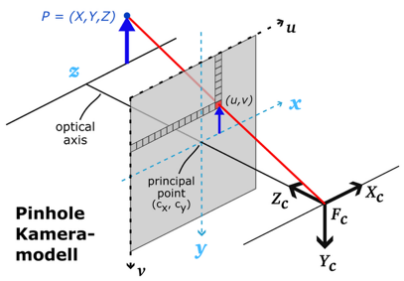
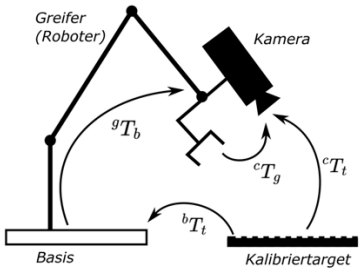
2 Wissenschaftlicher Hintergrund und Motivation

Aufgrund der Skalierbarkeit des Gesamtsystems ist eine zentrale Frage: "Wie kann man ein (möglichst) generisches System zum Tracking der Patientenposition aufbauen, dass die physischen Gegebenheiten des Behandlungssystems optimal abbildet und eine bewertbare, hohe Genauigkeit sicherstellt?" Bestehende Methoden aus OpenCV (Open Computer Vision Library), wie die Hand-Auge-Kalibrierung erfüllen zwar die Anforderungen, bieten aber mangels Skalierbarkeit und Erweiterbarkeit keine flexible Lösung, da die Geometrien fest im Code implementiert sind.

2.1 Einordnung

Mit dem OpenCV *Pinhole*-Kameramodell kann die Projektion von Objektpunkten (P_w) auf einen Sensor (P_c) modelliert werden (Tab. 1, oben). Basis ist ein physikalisch-mathematisches Abbildungsmodell inklusive typischer Abbildungsfehler von Objektiv-Sensorkombinationen [4, 5]. Es hat 6 Parameter der äußeren Orientierung ($^{w}R_c$, $^{wt}t_c$) und bis zu 13 Parameter der inneren Orientierung (f_x, f_y, c_x, c_y, K1, …). Nach Bestimmung dieser Parameter („Kamerakalibrierung"), können mit dem Modell Messung durchgeführt und 3D-Koordinaten berechnet werden.

Tab 1. Beschreibung der Hand-Auge-Kalibrierung aus OpenCV [4]

	Beschreibung	Transformation	
Kameramodell		Objekt auf Kamera: $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \cdot {}^{c}T_t \cdot \begin{bmatrix} X_t \\ Y_t \\ Z_t \end{bmatrix}$	Kameramatrix: $K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$
		Transformation Target in Kamera: ${}^{c}T_t = \begin{bmatrix} {}^{c}R_t & {}^{c}t_t \\ 0 & 1 \end{bmatrix}$	
Strukturmodell		${}^{c}T_t$ wird durch die Transformationskette ersetzt: ${}^{c}T_t = {}^{c}T_g \cdot {}^{g}T_b \cdot {}^{b}T_t$	
		Wobei: ${}^{c}T_g = \begin{bmatrix} {}^{c}R_g & {}^{c}t_g \\ 0 & 1 \end{bmatrix}$, ${}^{g}T_b = \begin{bmatrix} {}^{g}R_b & {}^{g}t_b \\ 0 & 1 \end{bmatrix}$, ${}^{b}T_t = \begin{bmatrix} {}^{b}R_t & {}^{b}t_t \\ 0 & 1 \end{bmatrix}$	

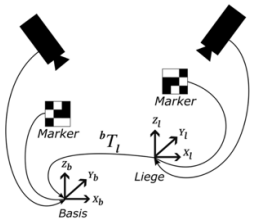
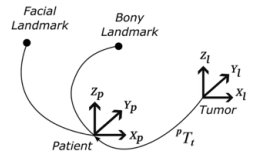
Im Beispiel der Hand-Auge-Kalibrierung wird, aufbauend auf das Kameramodell, die äußere Orientierung der Kamera ($^{c}T_g$ = [$^{c}R_g$, $^{c}t_g$], Kamera in Greifer) um die Transformationen des Roboters ($^{b}T_g$, Greifer in Basis) und des Kalibrierobjektes ($^{t}T_b$, Target in Basis) erweitert. Die äußere Orientierung dieses Gesamtaufbaus umfasst 18 Parameter. Die Pose des Greifers wird als Sollkoordinate angenommen. Insgesamt stellt dies einen geschlossenen

Vektorzug zwischen Objektpunkt und Bildpunkt über die zugrunde liegende Struktur des Roboters dar. Das Gesamtsystem muss kalibriert werden, bevor Messungen durchgeführt werden. Die Messungen erlauben dann die Bestimmung von Objektkoordinaten relativ zum Greifer, die in das Basiskoordinatensystem transformiert sind.
Wir haben ein Framework für die relative Verortung modularer Komponenten entwickelt, mit dem sich beliebig skalierbare Strukturen modellieren lassen. Aufbauend auf den gleichen Kamera- und Strukturmodellen werden Strukturdefinition, Ableitung und Handhabung der Modellparameter konsequent skalierbar umgesetzt.

2.2 Struktur des Gesamtsystems

Im Demonstrator für die liegenden Behandlung (Abb. 1, links oben) ist die Patientenliege um +/- 90° schwenkbar und der Bestrahler (Linearbeschleuniger, *LINAC*) rotierbar. Die Tracking-Kameras werden fest unter der Raumdecke montiert und blicken auf den Patienten, die Liege und den Raumboden. Die relevanten Objekte erhalten Marker zur zuverlässigen und sehr genauen Positionsbestimmung. Die Anlage zur halb-aufrechten Behandlung besitzt eine andere Objekteanordnung, führt aber auch eine Relativbewegung zwischen Liege und Bestrahler aus. Die Kameras werden ebenso im Raum montiert, können optional aber auch an bewegte Objekte angebracht werden.

Tab 2. Systemstruktur der liegenden Behandlung: links – vereinfachte schematische Darstellung, rechts – abgeleitete Objekte zum Tracking, wie sie aktuell umgesetzt sind

Komponente		Beschreibung
Anlage		Basis (Koordinatensystem) ○ Referenzmarker (255 Aruco-Marker, 6+2 Strecken) ○ 6 Kameras (ortsfest) und 1 Kamera (bewegt) ○ Liege (System, aktiv bewegt) ▪ Referenzmarker (Aruco 66, kreisförmig 54) ▪ 2 Kameras (fest)
Patient		Patient (Koordinatensystem) ○ Facial Landmarks (Version V1: 13, V3: 72) ○ Bony Landmarks (V1: 13) ○ Tumor (System)

Bei der liegenden Behandlung (Tab. 2) ermöglichen Marker im Raum (Basiskoordinatensystem) die Orientierung von Kamera und Liege. Die z.Z. verwendeten Aruco-Marker haben eine vergleichbare Unsicherheit und Genauigkeit wie Texturmerkmale, weshalb ein Umstieg auf kreisförmige Marker geplant ist. Die Patientenliege wird ebenfalls als Koordinatensystem modelliert, wobei die Drehachse im Raum (Basis) feststeht. Wir gehen davon aus, dass sich das mechanische System nicht nennenswert deformiert. Sechs fest installierte Tracking-Kameras überblicken den gesamten Behandlungsbereich und ermöglichen eine hoch genaue äußere Orientierung. Zudem wird zum Test verschiedener Blickwinkel eine frei bewegliche Kamera modelliert. Zur Verifikation der vorherigen Untersuchungen sind weiterhin drei Kameras auf der Liege installiert, die jedoch später kein Teil des Systems mehr darstellen.
Analog wird auch das Kopfmodell des Patienten aufgebaut. Aus dem Planungs-CT erhalten wir in einem gemeinsamen Koordinatensystem (Patient) die Lage und Orientierung des Tumors (Tumor), die Lage der knöchernen Merkmalspunkte (Bony-Landmarks, zur späteren Verifikation) sowie abgeleitet die Lage der Gesichts-Merkmalspunkte (Facial Landmarks, über RGB-Bilder).

3 Systemkalibrierung

Für das Kopf-Tracking muss das Gesamtsystem kalibriert werden. Wir haben uns für einen mehrstufigen Prozess, bestehend aus Kalibrierung der 3D-Koordinaten von Raum und Liege sowie der Kalibrierung der inneren und äußeren Orientierung der Kameras entschieden.

3.1 Kalibrierung der 3D-Koordinaten

Für die Anlage (Abb. 2 rechts) wurden mit einer hochauflösenden, handgeführten Kamera 71 Bilder aufgenommen. Der Raum (L=6100mm, B=3400mm und H=4200mm) wurde mit 255 Aruco-Marken versehen, die Liege mit je 66

Aruco-Markern und 54 kreisförmigen Markern. Weiterhin wurden im Raum 8 Strecken als Maßstab sowie zwei als Kontrollgröße mit einem Laserdistanzemesser bestimmt. Das Gerät besitzt jedoch eine Messgenauigkeit von nur ca. ± 2 mm und eine zusätzliche Unsicherheit der Antastung von ca. 1 mm. Die Kalibrierung wurde als klassische Bündelblockausgleichung (incl. Bestimmung der Inneren und Äußeren Kameraorientierung) durchgeführt, bei der 3D-Koordinaten, innere und äußere Orientierung der Kamera gemeinsam bestimmt wurden. Die Bildmessung fiel mit einem Residuum von σ_0 = 1,32 Pixel (Max: 8,23 Pixel) ungewohnt unsicher aus. Auch bei den Kontrollstrecken waren deutliche Abweichungen mit -3.92mm (3857mm) sowie -4.12mm (3868mm) erkennbar. Die Unsicherheit der 3D-Koordinaten ist in Tab. 3 für die einzelnen Objekte zusammengestellt.

Tab 3. Unsicherheiten der 3D-Koordinaten für Liege und Raum (Mittel- und Maximalwert)

	Mittelwert (Maximum) der Streckenunsicherheiten in mm			
	σ_x	σ_y	σ_z	σ_{xyz}
Liege, Aruco	0,03 (0,07)	0,05 (0,13)	0,12 (0,40)	**0,14** (0,42)
Liege, Kreisförmig	0,04 (0,08)	0,06 (0,13)	0,14 (0,42)	**0,15** (0,44)
Raum, Aruco	0,60 (1,62)	1,39 (4,46)	0,24 (1,43)	**1,65** (4,72)

Die mittlere Unsicherheit der 3D-Koordinaten der Liege sind mit ca. ± 0,15 mm auf einem annehmbaren Niveau. Jedoch zeigte sich mit ± 1,65 mm eine deutliche Unsicherheit bei den Raumkoordinaten. Ursache ist v.a. die schlechte Sicht auf die Deckenmarker, die für die hier verwendete Kameraanordnung nicht relevant sind. Weiterhin ist die Unsicherheit der kreisförmigen Marker auf der Liege in der gleichen Größenordnung wie die der Aruco-Marker, was unplausibel ist. Bei der Kontrolle wurde festgestellt, dass die Marker auf der Folie im Bildsatz aus schlechten Blickwinkeln aufgenommen wurden, sodass nur wenige Marker erkannt wurden und diese keine guten Triangulationsbedingungen hatten. Die Messung sollte also wiederholt werden.

3.2. Kalibrierung der feststehenden Kameras

Auf Basis der 3D-Kordinaten des Raumes sowie eines zusätzlichen Kalibrierobjektes, ist das Ziel des zweiten Schrittes, die innere und äußere Orientierung der feststehenden Kameras zu bestimmen. Das Kalibrierobjekt wird dabei, analog dem Vorgehen zur Kalibrierung von Motion-Capturing-Systemen, im relevanten Blickbereich bewegt und synchron Bilder aufgenommen. Insgesamt wurden dafür Bilder von 106 Aufnahmezeitpunkten mit ca. 25600 Bildmessungen aufgenommen. Das verbleibende Residuum fiel dabei mit 0,75 Pixeln (Max 6,99 Pixel) wesentlich geringer aus als im ersten Schritt. Es wurde ein Kameramodell mit einem Verzeichnungsansatz nach Brown [5] verwendet. Die Unsicherheit der äußeren Orientierung liegt im Bereich von wenigen mm (Lage). Für das Gesamtvorhaben ist das gerade noch ausreichend und unter Berücksichtigung der schlechten Bestimmung der 3D-Koordinaten des Raumes noch hinnehmbar. Die Unsicherheiten der inneren Orientierung sind auf dem ersten Blick akzeptabel. Die Unsicherheiten der Markerkoordinaten des Kalibrierobjekts, die bei der Kamerakalibrierung mitbestimmt werden können, sind mit ca. 0,09 mm (Tab. 6) besser als die äußere Orientierung erwarten lässt. Das bedeutet, dass der zentrale Bereich zur Überwachung der Kopfpose insgesamt gut kalibriert werden konnte.

Tab 4. Unsicherheiten der 3D-Koordinaten Kalibrierkreuz (Mittel und Maximum)

	Mittelwert (Maximum) der Unsicherheiten			
Arucos des Kalibrierobjekts	σ_x	σ_y	σ_z	3D
	0,05 (0,18)	0,03 (0,07)	0,06 (0,15)	**0,09** (0,18)

4. Bestimmung der 3D-Koordinaten der Facial Landmarks

Für das Patienten-Tracking erfolgt schließlich die Ableitung der 3D-Koordinaten der Facial Landmarks in einem separaten Prozess, da sie patientenspezifisch sind. Pro Patienten gibt es einen RGB-Bilddatensatz und einen Datensatz des Planungs-CTs/MRTs. Aus ersterem werden Facial-Landmarks mit 3D-Koordinnaten bestimmt. Dabei werden durch Mimik unveränderliche (z.B. Augen, Ohren) und veränderliche (z.B. Mund-, Kinnbereich) Landmarks

unterschieden (Abb. 3 links). Im CT/MRT-Planungsdatensatz werden Tumorposition, -orientierung und -kontur bestimmt. Zudem wird ein Referenzdatensatz von Facial- und Bony-Landmarks bestimmt (Abb. 4 rechts). Die Definition der Facial-Landmarks ist dabei identisch. Die Bony-Landmarks dienen zur Identifikation der Kopfpose anhand knöcherner Merkmale vor der Behandlung. Damit ist die Patientenposition stets durch zwei unabhängige Systeme sichergestellt. Die Merkmalsdefinition ist so gewählt, dass Facial- und Bony-Landmarks nah beieinander liegen und in Gesichtspartien mit dünner Hautauflage, so dass Hautveränderungen nur einen geringen Einfluss haben. Dabei wurde auf Markerdefinitionen der Forensik zurückgegriffen, die eine Rekonstruktion der Gesichtsform aus Knochenstrukturen erlaubt und größtenteils auch unseren Trackinganforderungen entsprechen.

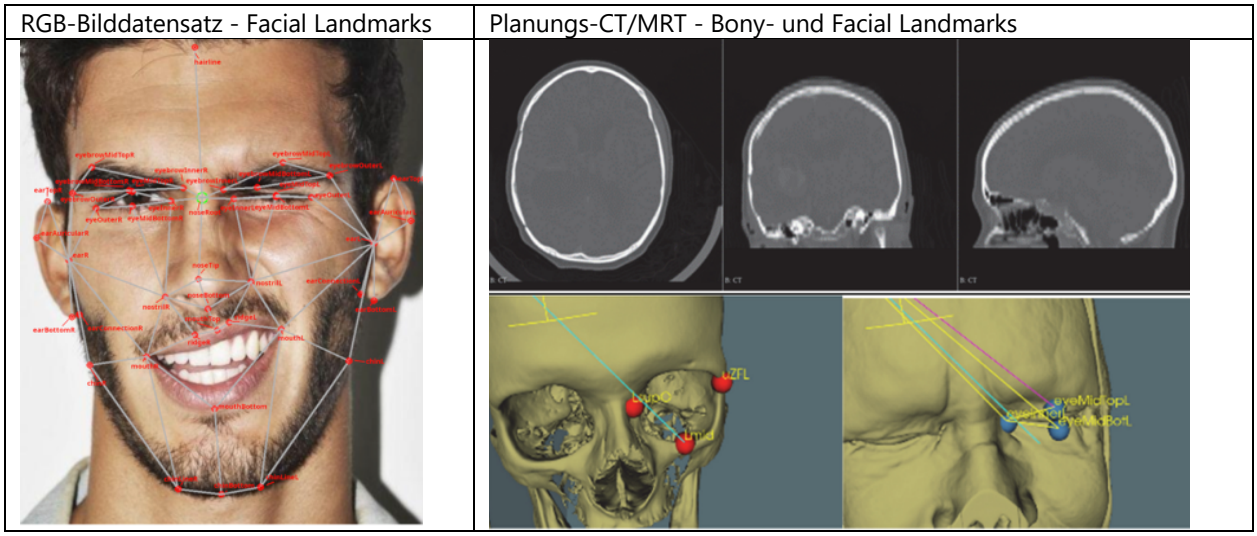


Abb. 3. Datensätze und Landmarks aus den Patientendatensätzen

Für die Detektion werden aktuell zwei 3D-CNN-Modelle aufgebaut, für die die Architektur der bestehenden RGB-basierten CNN-Modelle konsistent erweitert wurde. Aktuell wird das Training mit den in Abb. 3 dargestellten Landmarks aus unserem Annotierungs-Tool begonnen. Version 1 der Facial-Landmarks (14 Merkmale) wurde getestet. Tab. 5 zeigt die Unsicherheiten der aus 14 Bildern bestimmten 3D-Koordinaten. Mit durchschnittlich ± 0,14 mm ist die Unsicherheit größer als die der 3D-Markerkoordinaten (Tab. 4 und 5) aber akzeptabel. Die durch die Perücke des Puppenkopfes teils verdeckten Merkmalspunkte der Ohren zeigen die größten Unsicherheiten. Die 3D-Landmarkkoordinaten werden mit dem Planungs-CT/MRT abgeglichen, um Merkmalsveränderungen und Transformation auf den Tumor zu prüfen. Für die folgenden Schritte wurde das zunächst ausgelassen.

Tab 5. Unsicherheiten der Facial Landmarks (V1)

	Mittelwert (Maximum) der Unsicherheiten			
Facial Landmarks (14)	σ_x	σ_y	σ_z	3D
	0,06 (0,45)	0,05 (0,21)	0,11 (0,97)	**0,14** (1,09)

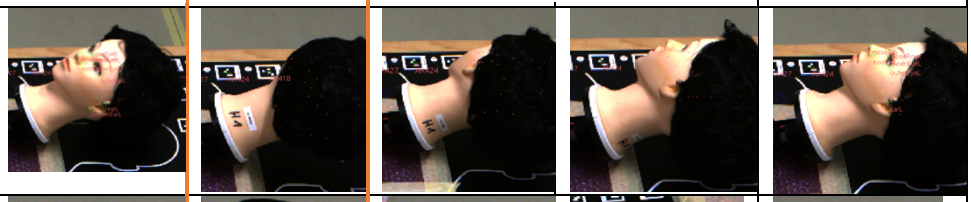
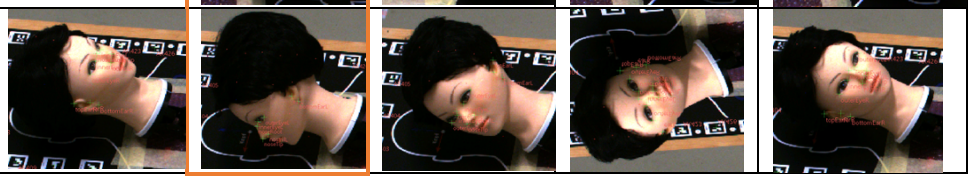
5. Tracking

Beim Tracking wird nun die Kopfpose jedes Aufnahmezeitpunkts relativ zur Liegenpose bestimmt. Die gewählte Modellstruktur ordnet den Kopf der Liege explizit zu, so dass die Unsicherheiten nicht wie in üblichen Modellen durch die Transformationen über Kopf in Raum sowie Liege in Raum implizit bestimmt werden. In Tab. 6 sind Unsicherheiten der für das Tracking relevanten Beziehung des Kopfes relativ zur Liege dargestellt. Auffallend ist die sehr geringe Unsicherheit der Liegenpose, die mit ca. ±0,02 mm (Lage) deutlich besser ist, als die Bestimmung der 3D-Koordinaten erwarten lässt.

Die Unsicherheiten der Kopfpose sind im ersten Aufnahmezeitpunkt mit ca. ± 0,4 mm in einem, für den aktuellen Bearbeitungsstand, akzeptablen Bereich. Verständlicherweise sinkt die Detektionsgüte bei extremen Blickwinkelbereiche (rot markierter Aufnahmezeitpunkt 2, ca. ± 5 mm). Relevant ist zu sehen, bis wohin die erforderliche Sicherheit noch erreicht werden kann. Bei großen, aber natürlichen Kopfbewegungen ist die

Unsicherheit bereits akzeptabel (± 0,5 mm, Zeitpunkte 3-5), was für die Stabilität der Landmarkdetektion spricht. Diese großen Blickwinkelbereiche erreichen die von uns getesteten kommerziellen Detektoren nicht.

Tab 6. Unsicherheiten von Liege und Kopfpose (V1) sowie beispielhafte Bilder der hochauflösenden Kameras (D1 und D2) um die durchaus extremen Blickwinkelunterschiede darstellen zu können

Head-Pose	X	+/- 1.893E-001	+/- 1.329E+000	+/- 4.671E-001	+/- 2.197E-001	+/- 2.150E-001
Head-Pose	Y	+/- 3.675E-001	+/- 3.935E+000	+/- 4.650E-001	+/- 1.772E-001	+/- 2.438E-001
Head-Pose	Z	+/- 4.203E-002	+/- 3.999E+000	+/- 8.647E-001	+/- 1.458E-001	+/- 7.523E-002
Head-Pose	U	+/- 5.553E-002	+/- 2.522E-001	+/- 8.331E-001	+/- 9.658E-002	+/- 6.368E-002
Head-Pose	V	+/- 8.234E-002	+/- 3.809E-001	+/- 8.668E-002	+/- 4.361E-002	+/- 6.654E-002
Head-Pose	W	+/- 2.535E-002	+/- 7.352E-001	+/- 7.743E-001	+/- 7.022E-002	+/- 3.697E-002
Kamera D1						
Kamera D2						

6 Zusammenfassung

Im Beitrag haben wir ein Framework für die relative Verortung modularer Komponenten entwickelt. Damit lässt sich ein beliebig erweiterbares System aus frei beweglichen Kameras und anderen Systemkomponenten modellieren. Die Komponenten werden mit optischen Markern vermessen, deren relative Lage bestimmt und über die Zeit präzise getrackt. Wir haben als Beispiel die Systemmodellierung einer Tumor-Bestrahlungsanlage vorgestellt. Die Verortung von Patienten erfolgt anhand von Gesichts- und knöchernen -merkmalen. Wir sind auf die konkrete Abbildung der physischen Beschaffenheit der Bestrahlungsanlage und auf deren Konsequenz für die reale Messung eingegangen. Weiterhin wurde die am Demonstrator erreichte Unsicherheit und Genauigkeit von Systemkalibrierung und Tracking an realitätsnahen Szenarien dargestellt. Wir schlussfolgern, dass der gewählte Systemansatz funktioniert. Auch wenn noch Optimierungspotenzial besteht, ist die geforderte Genauigkeit und Unsicherheit erreichbar.

7 Danksagung

Das Projekt „Intelligente Patientenpositionierung in der Präzisionsstrahlentherapie" ist in den Cluster 4 Future: „SEMECO" integriert und wird durch öffentliche Mittel gefördert.

CLUSTERS 4 FUTURE Next generation innovation networks

SPONSORED BY THE Federal Ministry of Education and Research

Kofinanziert von der Europäischen Union

Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des vom Sächsischen Landtag beschlossenen Haushaltes.

Referenzen

[1] J. Kreissl, „Entwicklung und Validierung eines Software-Frameworks zur Modellierung medizinischer Roboter", Diplomarbeit, Technische Universität Dresden, Dresden, 2024.

[2] M. Riedel, D. Schneider, und A. Bienert, „Photogrammetrische Bestimmung von Kopfbewegungen mittels Facial Landmarks – lessons learned", in GFal Tagungsband 2023, GFal, Nov. 2023.

[3] M. Riedel, D. Schneider, und J. Kramp, „Photogrammetrische Bestimmung von Kopfbewegungen mittels Facial Landmarks – proof of concept", gehalten auf der Oldenburger 3D-Tage, Feb. 2022.

[4] Dokumentation OpenCV Hand-Auge-Kalibrierung (angerufen am 26.08.2026) - https://docs.opencv.org/4.5.4/d9/d0c/group__calib3d.html#gaebfc1c9f7434196a374c382abf43439b

[5] T. Luhmann, S. Robson, S. Kyle, und J. Boehm, Close-range photogrammetry and 3D imaging. Walter de Gruyter GmbH & Co KG, 2023.

From Data-Driven to Physics-Informed: Benchmarking Neural Surrogate Architectures for Transient Thermal Prediction in Industrial Casting

Maben, Leonardo H.[1], Heidenreich, René[2]
[1]Gesellschaft zur Förderung angewandter Informatik e. V., Volmerstraße 3, 12489 Berlin, maben@gfai.de, www.gfai.de
[2]Gesellschaft zur Förderung angewandter Informatik e. V., Volmerstraße 3, 12489 Berlin, heidenreich@gfai.de, www.gfai.de

Abstract: Industrial casting simulations rely on computationally expensive Finite Element or Computational Fluid Dynamics methods, with runtimes ranging from hours to days depending on part complexity, which in turn limits the number of design variants that can practically be evaluated during the initial design process. This paper investigates a solution that can target this bottleneck, with the goal being the use of machine-learning surrogate models that can serve as complementary components to accelerate this workflow rather than being a replacement for high-fidelity simulation. The end result is a unified point-cloud representation across three structurally distinct geometries (cylinder, ring, L-shape), evaluated through a systematic four-way benchmark (residual MLP baseline, Physics-Informed Neural Network, PointNet, and Physics-Informed PointNet) carried out under identical training and evaluation conditions. Models were assessed on both predictive accuracy (RMSE, normalised RMSE) and inference throughput, with the results indicating that model suitability is geometry-dependent: the MLP and PINN achieve substantially higher inference throughput and strong accuracy on simpler geometries, but their error increases sharply on the more complex L-shape geometry, while PointNet and PIPN maintain stable accuracy across all three geometries at a higher computational cost. These findings suggest that no single architecture is optimal across all cases, and that model selection should instead balance geometric complexity against the accuracy–speed trade-off. The developed models are intended to serve as a fast add-on for preliminary configuration screening, with final designs still to be verified through conventional simulation tools.

Keywords: Surrogate Modelling; Physics-Informed Neural Networks; Point Cloud; Thermal Field Prediction

1 Introduction

Industrial casting simulations are a vital part of the systematic evaluation of the entire casting process. Over the last few decades, a workflow to simulate this process has been developed and is now seen as the norm in many industrial settings. This workflow usually consists of CAD models describing the object and process, which is accompanied by a numerical simulation through Finite Element Methods (FEM), Computational Fluid Dynamics (CFD) Methods or a combination of both the methods [1] which tend to be accurate. The challenge presented by these realistic casting simulations is the demand of modelling complex physics which results in computation times that can range from several hours to days depending on the shape and complexity of the casting object. Other hurdles include the cost of software licences and the vast expert knowledge required to utilise these tools.

The development and optimisation of metallic castings take place through iterative design of various sprue systems and feeder configurations. Due to volumetric shrinkage during solidification, risers must be carefully dimensioned to ensure adequate feeding and avoid defects such as porosity and shrinkage cavities. The thermal behaviour in sand casting is further influenced by the insulating properties of the sand mould, which significantly affects heat dissipation rates and solidification dynamics. This typically requires iterative geometry adjustments and repeated simulation cycles, each demanding significant computational resources and limiting the number of design variants that can be practically evaluated.

This paper investigates a solution that can target this bottleneck, with the goal being the use of machine learning models that can serve as complementary components that enable rapid thermal field approximation within exist-

ing casting development workflows rather than being a replacement for conventional simulation tools. The end result is a unified point cloud representation across three structurally distinct shape families (cylinders, ring, L-shape) evaluated through a systematic 4-way benchmark (MLP, PINN, PointNet, PIPN) under identical training/evaluation conditions. This paper is organised as follows. Section 2 presents a list of related works that have dealt with the various approaches to using physics-informed models. Section 3 consists of a description of the data used during training and descriptions of the different architectures benchmarked in this paper. In Section 4, the benchmarked results are tabulated and showcased. The final section contains a summary of the future work that can be done in this field to further improve performance.

2 Related Work

The following research builds on previous works in this field. The utilisation of Physics-Informed Neural Networks was first established through the work of Raissi [2] which set the basis for extending traditional MLP models to solve PDEs. This was foundational work with certain limitations such as the possibility of only one geometry per training, the need for retraining when geometry changes are present and geometry not being explicitly encoded. Building on the success of PINNs, Kashefi & Mukerji [3] proposed Physics-Informed PointNet (PIPN), which extends the physics-informed training paradigm into a PointNet architecture [4], thus addressing the geometric limitations present in the vanilla PINN model. PointNet improves local geometric feature extraction while considering point clouds. Through this architecture, irregular domains could be learnt and transferred to unseen geometries as well. PointNet-based models have been widely adopted in CAD processes.
Although surrogate models, PointNet-based geometric learning, PINNs and PIPNs have each demonstrated promising capabilities for thermal-field prediction, prior work has largely investigated these approaches independently. To the best of our knowledge, previous studies have concentrated on a 2D-based approach [5] but there is no study that showcases a 3D-based approach. This gap is addressed through a systematic comparison of purely data-driven, physics-informed, and geometry-aware architectures under identical training conditions for predicting transient thermal fields in metal casting processes. The purely data-driven Multi-Layer Perceptron (MLP) model acts as a baseline for the benchmark against which the performance of other models can be evaluated.

3 Dataset and Models

3.1 Data and Point Cloud representation

The study was carried out using simplified geometries to assess whether the proposed framework could work under controlled conditions. The complexity of these objects was increased in every subsequent iteration to evaluate its robustness and generalisability. This project was a joint research project with partners Access e.V., RWP GmbH, and Knowtion GmbH, with Access e.V. and RWP GmbH providing the input data through casting simulations.

The initial dataset used was a three-cylinder setup. This was a pure thermal simulation on a pure iron casting material with a simulation time of 200 seconds. It was carried out using the STAR-CCM+ simulation software [6]. There were 243 variations of the simulation object in total. Three variations in the initial casting temperature, three variations to the diameter of the upper cylinder as well as the lower cylinder, and three variations to the height of the upper cylinder and of the lower cylinder. Although the simulation was carried out for 200 seconds only 9 specific timestamps were provided as input data. This ultimately proved to be a pitfall, resulting in poor results due to sparsity of the dataset with respect to the timestamp. To counteract this problem, a second dataset was produced consisting of a pure thermal sand casting simulation of a two-cylinder object made of a cast iron material, GJS-400-18-LT. Due to limitations on how long the simulation would take to run to solidification only 10 variations were carried out, with the varying parameter being the diameter of the lower cylinder. Each of the variations contained 81 timestamps for a simulation run of 20k seconds (roughly 5.5 hours). This was ultimately chosen as the first viable dataset with which

to carry out tests. Further datasets were provided by RWP using the WinCast® simulation software [7]. These were a ring object with 10 total variations, two distinct starting temperatures and 5 different feeder positions for a total runtime of 250 seconds. In this iteration we received all 250 timestamps and used a similar configuration to produce an additional dataset on an L-shape feeder system with 9 variations of feeder positions with a total simulation runtime of 840 seconds.

The dataset is primarily a time series of 3D point clouds and initially consists of a variety of physical properties such as, temperature, heat transfer coefficient, cooling rate, temperature gradient, solidification time, solid volume fraction, liquid residence time, Niyama criterion, pore volume fraction, and the volume fractions of liquid and voids. Most of these features are derived physical quantities. Porosity-related indicators such as the Niyama criterion, for instance, are computed by the project partner Access e.V. directly from the predicted temperature field in a downstream step, which is why temperature was retained as the primary target.
The data acquisition is followed by a pre-processing step. Pre-processing involved merging geometry and process level metadata such as, casting temperature and region identifier, onto each point by simulation ID. Outward surface normals at boundary points were estimated per unique mesh location, oriented away from the geometry centroid, for use in a boundary-condition loss term. The data was standardized to ensure consistent unit scales for the various features. The data was split into training, validation and test sets at simulation level to prevent leakage between the datasets. Through pre-processing and determining which properties are only present during inference, the final feature list is condensed. The final feature list represents each point x_i in this dataset and can be described through the following mathematical representation:

$$x_i = [X_i, Y_i, Z_i, t_i, T_{casting}, r_i], \tag{1}$$

where X_i, Y_i, Z_i are the coordinates in the point cloud, t_i is the timestamp, $Tcasting$ is the starting temperature and r_i is the region identifier. The target parameter $\hat{T}_i$ is temperature and can be represented through the following equation:

$$\hat{T}_i = T_\theta(x_i) = T_\theta(X_i, Y_i, Z_i, t_i, T_{casting}, r_i) \tag{2}$$

The neural network, parametrised by its learned weights and biases θ, predicts temperature from $X, Y, Z, t, T_{casting}, r_i$.
In addition to the simulation point cloud data, temperature dependent material properties were provided for the casting material and mould material. These properties are not treated as point-wise input features but are used to formulate the governing physical constraints incorporated into the physics loss.

3.2 Architectures

During the course of this research, four essential architectures were used: MLP, PINN, PointNet and PIPN.
MLP: This model serves as our baseline, a residual Multi Layer Perceptron. Huber loss is the metric used for penalising loss terms during training due to the dataset's broad temperature ranges (temperatures can go as high as 1800 Kelvin or as low as 800 Kelvin). The architecture used 6 residual blocks, Layer Normalisation, a width of 512 and a SiLU activation while also using a dropout rate of 0.3 to ensure overfitting is kept to a minimum through this regularisation technique.
PINN: A Physics-Informed Neural Network is in essence an extension of the MLP network. It is a residual fully-connected network and has an identical architecture to the MLP network above. PINN operates point-wise so there is no cross point/permutation mechanism. The theory being that spatial reasoning is expected to come from the physics loss, not the architecture. In simple terms, PINN utilises the standard loss function used in the MLP network with an additional physics loss which is described in section 3.3
PointNet: Unlike the point-wise MLP and PINN, this architecture reasons over the entire point cloud jointly. Per-point features are extracted through a series of shared Conv1D layers, with a T-Net module aligning both the input coordinates and the intermediate feature space. These per-point features are aggregated via max-pooling into a single global shape descriptor, which is concatenated back onto each point's local features before the final prediction. This allows the network to capture local structure that PINN and MLP cannot represent by construction.

PIPN: Essentially utilises the same extension applied in PINNs, to deliver a Physics Informed model that is theoretically the most expressive and also the most computationally expensive variant.

3.3 Physics Loss

Physics loss is the incorporation of physical laws described by differential equations into the loss functions of neural networks [2]. The intention behind this extension of the loss function is to guide the learning process toward solutions that are more consistent with physical laws. For this particular problem, the transient energy conservation equation is the governing equation [8].

$$\frac{\partial}{\partial t}\left[\rho(T)\left(c_p(T)T + LF(T)\right)\right] = \nabla \cdot \left(k(T)\nabla T\right) \tag{3}$$

It accounts for the temporal variation of the stored thermal energy, including both sensible and latent heat contributions, and the spatial redistribution of heat through thermal conduction. The thermal energy term is represented by the LHS of the equation, where ρ is the density, c_p is the specific heat capacity, *L* is the latent heat, and *F(T)* denotes the solid phase fraction, increasing from 0 (fully liquid) to 1 (fully solid) as temperature decreases. The heat conduction term is represented by the RHS, where *k(T)* is the temperature-dependent thermal conductivity. The above mentioned terms are derived through spline functions from the material data for GJS-400-18-LT provided for these simulations.

The PDE residual of this equation is then penalized and used as a physics loss term:

$$R_T = \left[\frac{d\rho}{dT}\left(c_p T + LF\right) + \rho\left(c_p + T\frac{dc_p}{dT} - L\frac{dF}{dT}\right)\right]\frac{\partial T}{\partial t} - \left[k\nabla^2 T + \frac{dk}{dT}\left|\nabla T\right|^2\right] \tag{4}$$

where R_T is the PDE residual. The Physics loss is ultimately the mean squared PDE residual over the collocation points:

$$\mathcal{L}_{\text{physics}} = \frac{1}{N_f}\sum_{i=1}^{N_f}\left|R_T(\mathbf{x}_i, t_i)\right|^2 \tag{5}$$

where N_f represents the total number of collocation points used to calculate the PDE, i is the index of these points, x_i is the spatial coordinates of the *i*-th collocation point and t_i describes the time coordinate.

For datasets spanning both a casting and a mould region, the residual is instead evaluated per region using that region's own temperature-dependent material properties, with each collocation point assigned to a region via an associated region label. Since the two regions can produce residuals of very different magnitudes, each region's residual is normalised by an exponential moving average of that region's residual standard deviation before being squared and averaged. This prevents whichever region has the larger residual magnitude from dominating the loss. The final multi-region physics loss is the unweighted average of the two normalised terms:

$$\mathcal{L}_{\text{physics}}^{\text{multi}} = \frac{1}{2}\sum_{r\in\{\text{casting},\ \text{mould}\}}\frac{1}{N_r}\sum_{i=1}^{N_r}\left(\frac{R_{T,r}(\mathbf{x}_i, t_i)}{s_r}\right)^2 \tag{6}$$

where N_r is the number of collocation points assigned to region r and s_r is the exponential-moving-average estimate of that region's residual standard deviation.

Along with the physics loss, there is a boundary condition which also needs to be considered, described at the mould boundary using a convective (Robin) condition:

$$-k(T)\,\frac{\partial T}{\partial n} = h\,(T - T_{amb}), \tag{7}$$

with $\partial T/\partial n$ from automatic differentiation, n estimated per point, $k(T)$ from a fitted material spline, and h learned jointly with the network. The resulting normalised residual loss $\mathcal{L}_{bc}$ enters the total loss. The overall loss term $\mathcal{L}$ is finally represented as the total sum of data, physics and boundary losses:

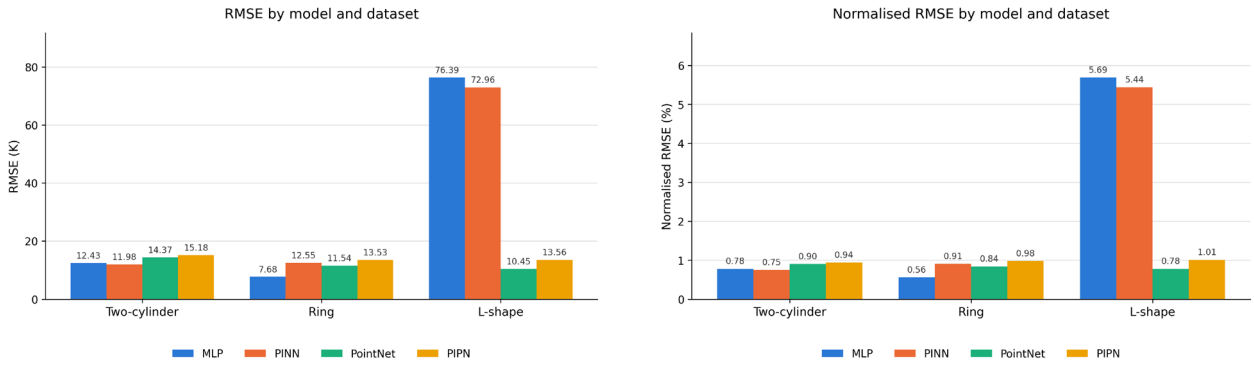
$$\mathcal{L} = \mathcal{L}_{data} + \lambda_{physics}\,\mathcal{L}_{physics} + \lambda_{bc}\,\mathcal{L}_{bc}. \tag{8}$$

3.4 Training protocol

The execution of this benchmark requires consistent training requirements to ensure viable metrics for comparison. Keeping this in mind, all models were trained on the same datasets with the same split. The dataset was split into training, validation and test sets, with 80% of the data used for training, 10% for validation and 10% for final evaluation. The non physics models were evaluated using a single configuration, whereas the physics based models were evaluated using three different values of the physics loss weighting parameter, $\lambda_{physics}$ $(1e-4, 1e-3, 5e-3)$. Since no established default value for $\lambda_{physics}$ exists, these values were selected to investigate the influence of the relative weighting of the physics-based component. For the non-physics models, no additional hyperparameter related to the physics loss is required, and therefore only a single configuration was evaluated. The best performing physics-based configuration was selected based on the validation set, while the non-physics models were evaluated using their fixed configuration. The test set was kept strictly separate and used only for the final performance comparison. It should be noted that the reported benchmarks are based on a single held-out test geometry per dataset; the next phase of the project should therefore evaluate the models on a larger number of samples to obtain more statistically robust performance estimates. All models were trained on the same hardware (RTX 4090) with multiple workers for the DataLoader class to ensure maximum use of the GPU during data handoff between CPU and GPU. The metrics used for evaluating the performance of the models are RMSE, normalised RMSE(% over a temperature range) and region-wise error evaluation.

4 Results

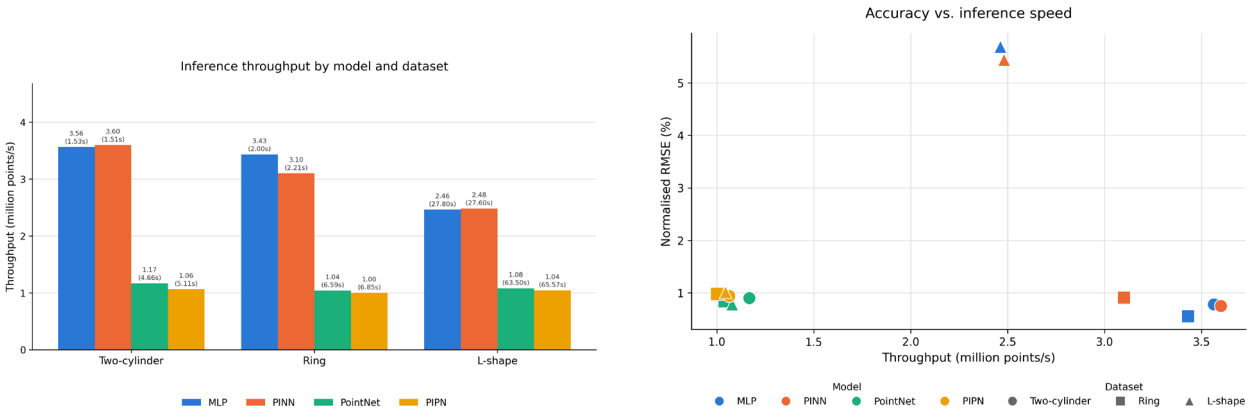
The RMSE chart shows absolute temperature prediction error in Kelvin for each model across the three test geometries. On two-cylinder and ring, all four models perform comparably (roughly 7.7–15.2 K), but on the L-shape casting dataset, MLP and PINN errors jumps sharply to 76.4 K and 73.0 K, while PointNet and PIPN stay in the same range as the other two datasets (10.5 K and 13.6 K)



(a) RMSE (b) Normalised RMSE

Fig. 1: Benchmark of ML models: (a) RMSE and (b) normalised RMSE across the two-cylinder, ring, and L-shape datasets.

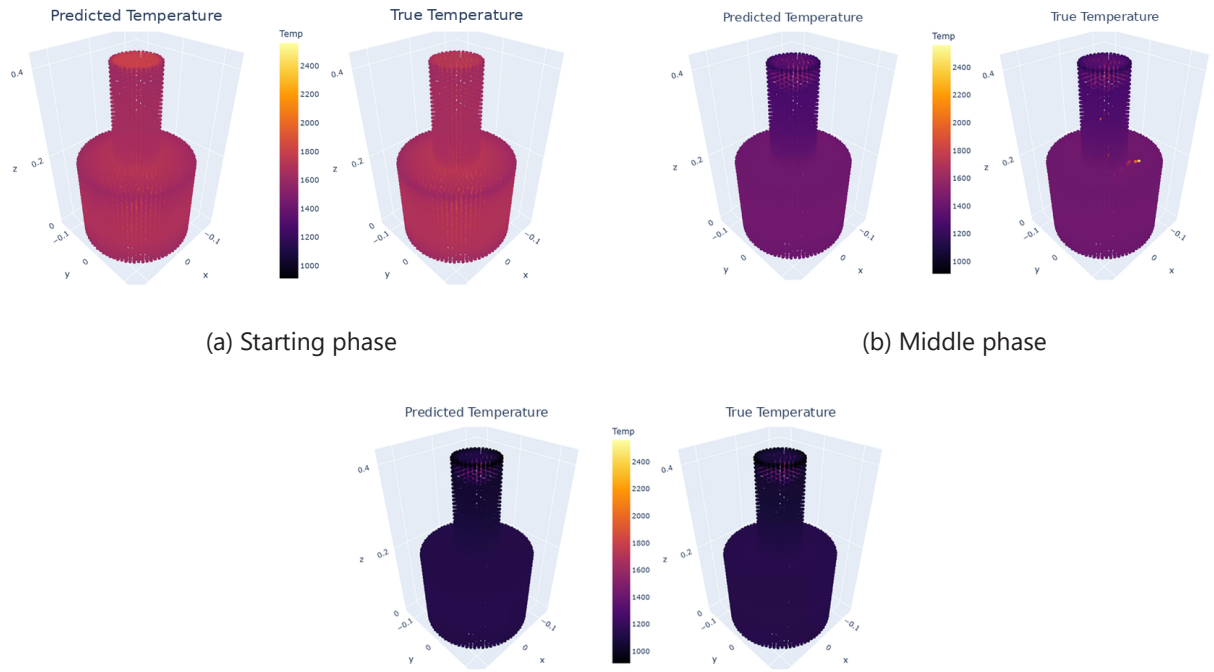
The normalised RMSE chart expresses the same error as a percentage of each dataset's temperature range, making the three geometries directly comparable on one scale: all models stay below 1% on two-cylinder and ring, but on the L-shape dataset MLP and PINN rise to 5.69% and 5.44% (five to seven times higher than on the other geometries) while PointNet and PIPN remain low at 0.78% and 1.01%, confirming that point-based models generalize substantially better to complex, non-convex geometry.



(a) Inference throughput (b) Accuracy vs Inference speed

Fig. 2: Model performance at Inference

Fig. 2 showcases the performance of the models during inference. Fig. 2a) contains both the throughput of each model as well as the total time needed for inference. Depending on the two-cylinder, ring, or L-shape dataset, the prediction times were at max roughly 5, 7, or 66 seconds respectively. MLP and PINN tend to be much faster in predicting temperature values for all three datasets. two-cylinder and ring objects were relatively simple and hence allowed for high throughput in general. The L-shape object was more complex while also being larger. PointNet and PIPN were slower in comparison as expected. However, when we look at the accuracy of the models as seen in Fig. 2 b), the best case scenario is to have values lying on the bottom right of the graph, i.e. a high throughput while also containing a low NRMSE value. For the simpler objects, MLP and PINN performed the best, however the error rate shot up with the L-shape object. PointNet and PIPN, however, provided stable results for all datasets, without the sharp degradation seen in MLP and PINN, although it might not be the fastest model at inference.



(a) Starting phase (b) Middle phase

(c) End phase

Fig. 3: Comparison of 3D temperature fields from the surrogate model and numerical simulation based on point clouds at three stages of the casting process: (a) Starting phase, (b) Middle phase, and (c) End phase.

Above is a point cloud representation of the PINN model evaluated on the two-cylinder dataset at three different stages of the simulation process. Temperature is colour-coded to show its magnitude at each stage. The predictions

closely match the true temperature values at all three stages, with some localized error concentrated in the middle stages of the simulation. This suggests the predictions are accurate enough to serve as fast approximations for quick configuration checks during the early stages of setup.

5 Acknowledgements

We thank our project partners Access e.V., RWP GmbH, and Knowtion GmbH for their trusted and constructive collaboration. Their expertise, the provision of data and many in depth discussions regarding the simulation software and sandcasting processes contributed significantly to this work.

6 Conclusion

Overall, this study demonstrates the potential of machine-learning surrogate models as a supporting tool for numerical simulation workflows in casting applications. The objective is not to replace existing simulation solutions, but to provide rapid temperature-field estimations that can assist engineers during early configuration and parameter-selection stages, where many different configurations may need to be explored. MLP and PINN offered substantially higher throughput across all geometries, making them the faster choice where applicable. However, with complex structures, the accuracy degraded considerably whereas PointNet's stable accuracy justified its high inference cost. The results therefore show that the most suitable model depends on the geometry, balancing accuracy against inference speed. Future work should therefore focus on evaluating the models across a larger number of test samples to obtain more robust performance estimates, improving generalisation to unseen geometries, optimising the physics-loss weighting, and exploring whether a hybrid or selection strategy could combine the speed of MLP/PINN with the robustness of PointNet across varied geometries. This could enable the developed models to serve as a fast add-on to existing simulation workflows, reducing the time required for preliminary configuration studies while retaining the numerical simulation as the final high-fidelity evaluation.

References

[1] J. S. S. Vajna, C. Weber and D. Schlottmann, *CAD/CAM für Ingenieure: Hardware, Software, Strategien*. Wiesbaden: Vieweg+Teubner Verlag, 1994.

[2] M. Raissi, P. Perdikaris, N. Ahmadi, and G. E. Karniadakis, "Physics-informed neural networks and extensions," 2024.

[3] A. Kashefi and T. Mukerji, "Physics-informed pointnet: A deep learning solver for steady-state incompressible flows and thermal fields on multiple sets of irregular geometries," *Journal of Computational Physics*, vol. 468, p. 111510, Nov. 2022.

[4] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," 2017.

[5] J.-w. Kang, J.-x. Zhu, and Q.-c. Zhao, "Temperature fields prediction for the casting process by a conditional diffusion model," *China Foundry*, vol. 22, no. 2, pp. 139–150, 2025.

[6] Siemens, "Simcenter star-ccm+ software."

[7] RWP, "Wincast expert finite element simulation software."

[8] V. R. Voller and C. Prakash, "A fixed grid numerical modelling methodology for convection-diffusion mushy region phase-change problems," *International journal of heat and mass transfer*, vol. 30, no. 8, pp. 1709–1719, 1987.

Additively Manufactured Sheet Metal Tools for High-Strength Steel Processing

Robin Roj[1], Johannes Weis[2], Mehrdad Mehrbakhsh[1], Benjamin Schelnberger[1], Christian Pelshenke[1], Jens Kneifel[1]
[1]Forschungsgemeinschaft Werkzeuge und Werkstoffe e. V.,
Papenberger Straße 49, 42859 Remscheid, roj@fgw.de, www.fgw.de
[2]Fraunhofer-Institut für Produktionstechnologie IPT, Verkettete Produktionsmaschinen,
Steinbachstraße 17, 52074 Aachen, johannes.weis@ipt.fraunhofer.de, www.ipt.fraunhofer.de

Abstract: This research aimed to make thermal assisted forming of high-strength steels significantly more efficient and to consider tool-integrated hardening in order to enable the economical use of thermal assisted sheet metal processing. In order to achieve these goals, structural and thermal optimization of the tools was necessary. Thus, a hybrid tool system for drawing dies with internally cooled active parts was designed and implemented for the first time, combining conventionally manufactured tool bodies with additively manufactured active elements containing integrated cooling channels. The approach was validated on three representative applications. In the first, a helical and a fountain-shaped cooling channel structure were integrated into a round punch and compared with respect to their thermal performance. In addition, an elongated drawing punch and a tool for the manufacture of bipolar plates were investigated as further application examples, each representing a different set of geometric and thermal boundary conditions. Due to the hybrid design of the three tooling systems and the integration of contour-adapted cooling channels, a direct comparison with the conventional tool variants used in series production showed that the tool temperature could be reduced in all three example processes. This lower thermal load had a positive effect both on the wear behavior of the tools and on the resulting workpiece quality. The results are particularly beneficial for the sheet metal processing industry, which increasingly has to work with high-strength materials.

Keywords: Additive Manufacturing; Bipolar Plates; High-Speed Steel; Laser Powder Bed Fusion; Sheet Metal Processing

1 Scientific Background and Motivation

In the sheet metal processing industry, there is a trend towards processing ultra-high-strength steels, as well as towards higher quality standards and tighter tolerances. As a result of this development, conventional sheet metal working processes are increasingly reaching their economic and technological limits due to high tool loads. During the shearing and forming of high-strength sheet materials or the high-precision machining of, e.g., bipolar plates for fuel cells, thermal fluctuations play a significant role, particularly in semi-hot or hot working.
To improve the formability of high-performance steels, thermally assisted sheet metal forming processes were developed and implemented. In these processes, the high-strength steel sheet is locally heated immediately before the forming process, thereby reducing its strength. The increased formability of the sheet material leads, e.g., compared to conventional cold shear cutting, to high clean-cut rates of up to 100% and to reduced process forces by up to 70%; during the forming, the expansion ratio can be increased by up to 100%. In addition, components can be locally work-hardened. Due to the higher machining temperatures, material buildup on the tool occurs more frequently during thermally assisted sheet metal processing, leading to an unacceptable reduction in the edge quality of the final product. Furthermore, the thermal expansion of the active components plays a significant role in terms of component tolerances. One basic idea for addressing these challenges is to counteract them using additively manufactured tools equipped with internal cooling channels and thereby opening additional application areas in the field of tool-integrated hardening. Furthermore, an adaptation of the tool materials is necessary to withstand high thermal and mechanical stresses on the cutting edge. Here, the potential of additive manufacturing lays in leveraging new opportunities to use innovative materials with ductile core and hard outer layer properties.

Collar forming tests characterize the difference between heated and unheated sheet metal processing. Here, the limit expansion ratio for collar forming of a dual-phase steel (DP1000) determines a reference process. The expansion ratio is defined as the ratio of the final diameter to the pre-punch diameter. The diameter of a standard tool is 20 mm. To optimize the limit expansion ratio, the sheet metal is locally softened around the pre-punch hole. Induction heating is used as the heating method. In this process, a magnetic field heats the sheet metal in a defined zone. Visible cracks and necking at the collar edge are considered failure criteria.
Unheated forming shows significant cracking at the collar at an expansion ratio of 2. This results in a pre-hole diameter of 10 mm, which is punched with a cutting gap of 0.2 mm. Beside the depiction of the machine for sheet metal processing, Fig. 1 illustrates the difference between a preheated sheet and a cold sheet (made of 1.4301 with a thickness of 3 mm). By locally heating the sheet metal to 800 °C, the forming force can be significantly reduced and the collar can be formed without defects.

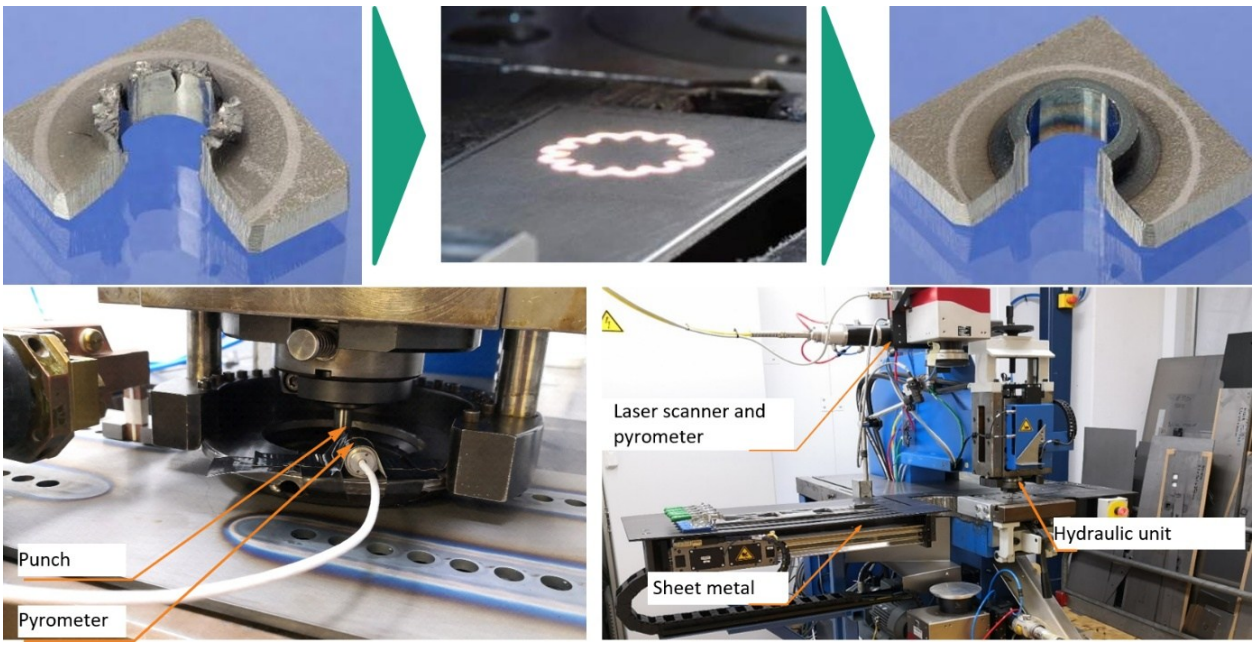


Fig. 1. Difference between cold and preheated sheet metal processing

Currently, conventional components for dies are manufactured using traditional abrasive or machining processes. These components are typically machined from a solid block; in some cases, a segmented design is used. After machining, a hardening and grinding process follows to achieve the desired cutting edge preparation (surface hardness, edge rounding, etc.). Optional polishing or coated surfaces increase the service life. However, diamond-like carbon and carbide coatings are used in particular for components subject to thermal stress and wear [1].
Additive manufacturing is increasingly becoming established for the production of active parts. Laser powder bed fusion (LPBF) is already being used industrially, particularly for complex geometries and topology-optimized dies [2]. Process optimizations in additive manufacturing are currently still the subject of research. The publications of Bidare et al. as well as Asnafi provide an overview [3, 4]. Regarding cooled tools, these are already being used in sheet metal processing in the areas of press hardening or hot bulk forming. Due to the large tool dimensions, cooling channels close to the contour can be incorporated relatively easily in these applications. Further investigations into the applicability of additively manufactured active parts for forging processes in the context of cooling requirements are evaluated, for example, by Chantzis et al. [5]. In addition, various construction strategies are employed. The active components can be manufactured entirely using a generative approach or built onto an existing base tool (hybrid construction method) [6].
As part of tool development for active components with internal cooling channels, another application scenario is relevant. The thermally assisted forming of bipolar plates for fuel cells or electrolyzes offers process-related advantages that push the boundaries of formability. Further process optimizations are emerging, particularly in the

context of the growing trend toward sheet metal components for hydrogen systems. Metallic bipolar plates are flat sheet metal components into which delicate flow channels for the reaction media are formed using hollow embossing. By locally heating the sheet metal prior to processing, the formability can be increased and the flow properties improved. This allows deeper flow channels to be formed without cracking, thereby optimizing the hydrogen mass flow from the bipolar plates. Further information on the manufacturing of bipolar plates is shown in Fig. 2.

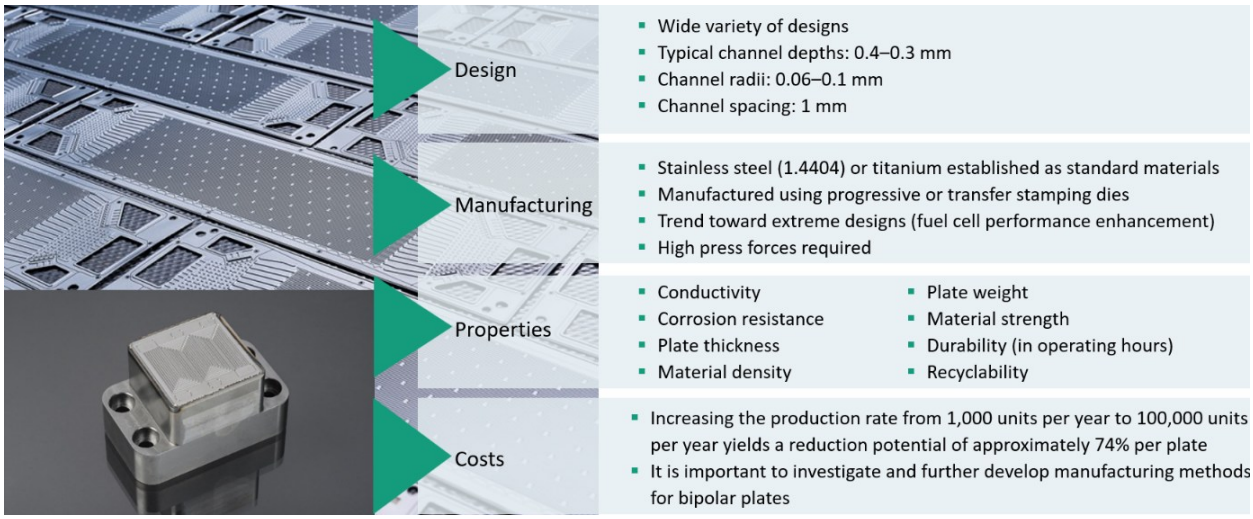


Fig. 2. Depiction and additional information about bipolar plates [7]

2 Materials and Methods

For the process of thermally assisted collar forming with internal cooling channels, it was necessary to define a die geometry that allows for the separation of the machine interface and the active die zone. Fig. 3 illustrates the upper die section, which acts as a machine adapter, as well as the lower, additively manufactured section with cooling channels. The adapter contains connections for the coolant as well as an internal feed to the active part. It was planned to manufacture the adapter conventionally for this purpose. A radially centered and flat-ground surface serves as the interface, with the connection realized by means of three screws. The additively manufactured active part can be customized as desired and equipped with individual channel configurations. Furthermore, this concept protects the machine from the imprecise interfaces of a fully additively manufactured punch and also reduces the volumes to be printed. This enables more cost-effective production of the complete tool, consisting of the punch adapter and the active part. The figure also illustrates the wear mechanisms of the punch and the sheet metal.

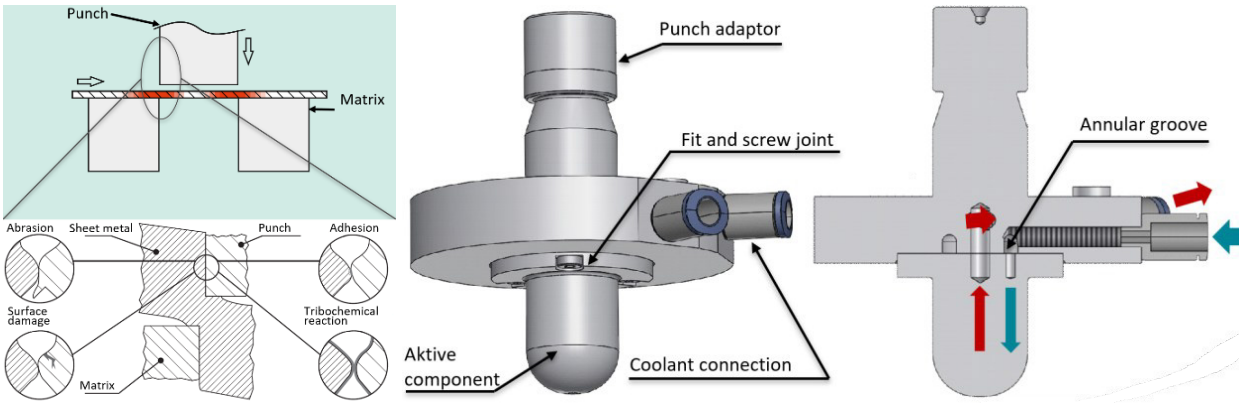


Fig. 3. Visualization of wear and tear as well as hybrid design for sheet metal processing

In order to systematically develop definitions for manufacturing processes, it was decided to test three exemplary tool shapes, produce the corresponding designs, put them into operation, and evaluate them (cf. Fig. 4).

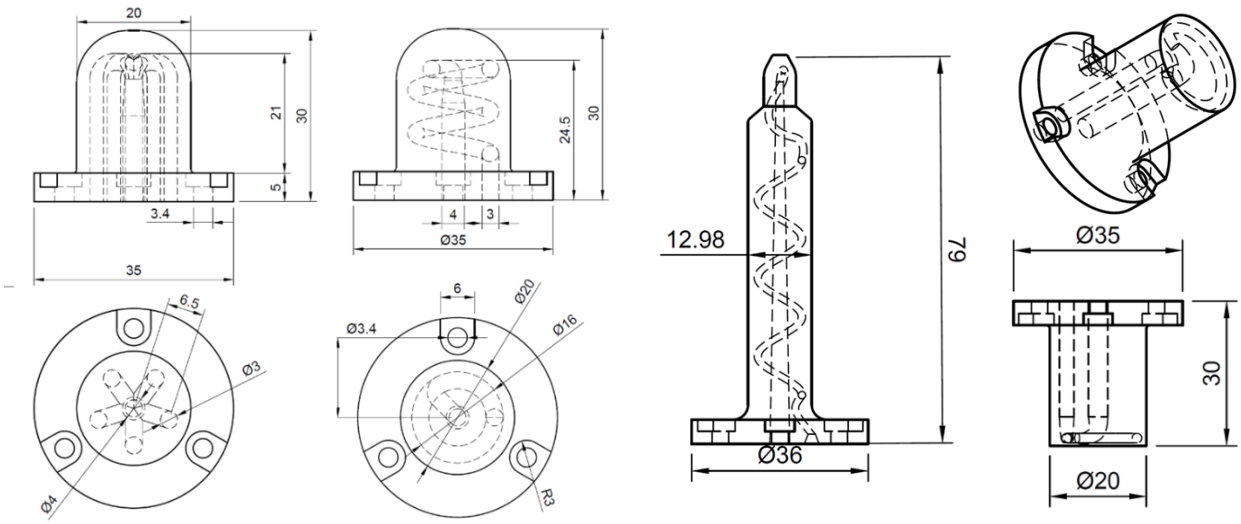


Fig. 4. Technical drawings of three different sheet metal tools

For the additive manufacturing of all components with integrated cooling channels, a 3D printer of the type MIDI from the company Aconity has been used to process high-speed steel PM23. To produce all other components that could be manufactured conventionally, milling processes have been used. In particular, the following use cases were developed: a round punch with fountain- and helix-type cooling, an elongated punch, and a tool for manufacturing bipolar plates. These were designed as modular systems, each consisting of the conventionally manufactured adapter for clamping into the machine and an additively manufactured component with a contour-optimized network of cooling channels. All three examples considered compatibility with the sheet metal processing machine as well as the use of a coolant. To implement the systems and determine the performance of all variants, an ECCO-Line EL 500 punching machine from Boschert as well as a Schuler MSC 2000 servo press were used.

3 Results

As part of the steps toward the additive manufacturing of active components, the LPBF process was shown to produce high-quality results with crack-free and low-porosity components. For the contour-optimized design of the cooling channels in all three examples, various CAE tools were successfully employed to conceptualize the hybrid tools and manufacture them using both additive and conventional processes (cf. Fig. 5). Both the design with different cooling channel networks and the manufacturing process followed by depowdering proceeded without issues. While the additive processing of PM23 proved to be challenging due to the high carbon content, manufacturing of high-quality, crack-free specimens for all three application examples using LPBF ultimately succeeded. The subsequent rework steps, involving machining of functional surfaces and targeted heat treatment, were also carried out in such a way that a sufficient number of hybrid demonstrators for all three application examples could be manufactured and tested. Due to the hybrid design of the three tooling systems and the incorporation of contour-optimized cooling, it was demonstrated, in a direct comparison with conventional variants in series production, that the tool temperature could be kept lower in all three example processes. This had a positive effect on both the wear behavior of the tools and the quality of the workpieces. The cooling performance of the AM collar-forming punches was experimentally investigated in a single-stroke test. A process-related advantage of a cooled active part results from reduced overheating of the active part during semi-hot and hot forming due to the cyclic contact between the active part and the heated sheet metal surface. For validation, the die surface temperature was measured with a pyrometer after the return stroke from the hot-formed collar. The setup and results are also shown in Fig. 5. A pyrometer was attached to the hold-down device to measure the temperature of the punch's outer surface at specific points. The temperature profile shown in the results diagrams can generally be divided into four sections. First, the punch temperature was measured before the stroke, and during the advance, the reflection from the heated

sheet metal on the punch surface was measured (advance phase). The drop in temperature after the feed stroke could be attributed to the punch movement, as the tool clamping mechanism moved in front of the pyrometer and shifted the focal point. After the punch's return stroke, the punch temperature following forming could be measured. The punch temperature after forming a collar with an uncooled AM punch was 28.5 °C. Cooling water at 18 °C was circulated through the AM punches, and the temperature of the helix punch after cooling was 25.3 °C.

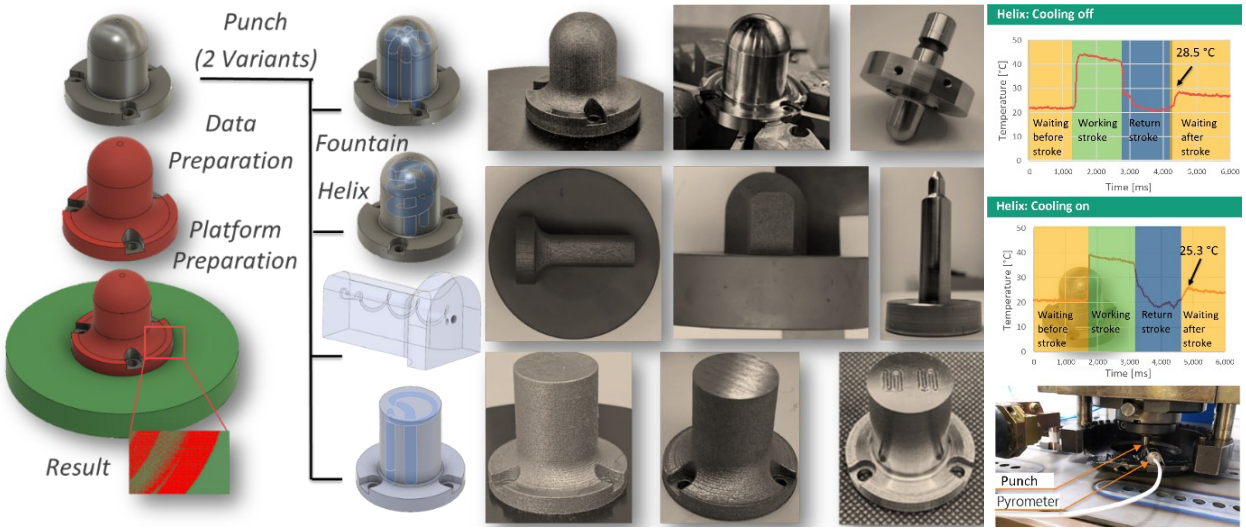


Fig. 5. Visualization of the process chain and additively manufactured specimens of all three example tools

The goal was to use the same tool adapter as for the collar forming tests. The active part was to be produced using additive manufacturing, and a typical channel structure found on bipolar plates was to be created using micro-milling. Mechanical strength was a particular consideration here. Using the punch and die tool pair, two serpentine-shaped channel structures can be formed (cf. Fig. 6).

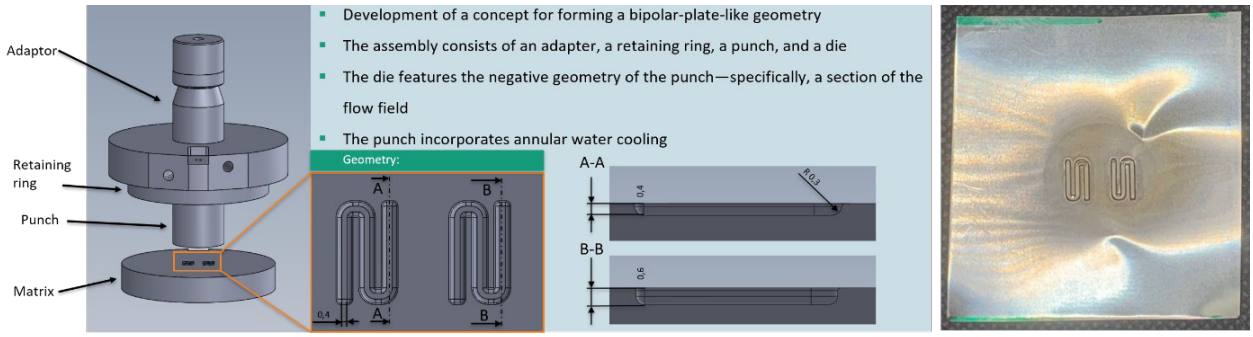


Fig. 6. Concept and result of the manufacturing of structures typical for bipolar plates

4 Discussion and Conclusions

The results achieved in this work demonstrate the economic potential that can be unlocked through the use of additively manufactured tools in thermal assisted sheet metal processing. The processing of ultra-high-strength steels remains technically demanding: the elevated process temperatures, the high contact pressures, and the resulting thermal and mechanical loads acting on the active tool elements place considerable demands on both tool design and tool material selection. Despite these challenges, it has been shown that the combination of additive manufacturing, hybrid tool concepts, and contour-optimized cooling does not merely yield advantages on a technical level - such as reduced tool temperatures, more stable process conditions, and more favorable wear behavior - but also translates into measurable improvements from an economic perspective. The decisive factor here is that the additional expenditure associated with additive manufacturing is offset by longer tool lives, more consistent component quality, and the possibility of realizing process chains that would otherwise require additional operations

or separate tooling. Against this background, the question of practical relevance is the extent to which these findings actually enable companies to make their production processes more efficient, to reduce their manufacturing costs, and to strengthen their market position in a sustainable manner. This applies in particular to small and medium-sized enterprises, for which investment decisions in new tooling technologies carry proportionally greater weight and for which the transferability of the results to their own product portfolio is a prerequisite for adoption. The present work provides the technological basis for such an assessment and thereby forms a starting point for the industrial implementation of hybrid, additively manufactured tool systems.

Acknowledgement

The research project IGF 23013 N (01IF23013N) "Additiv gefertigte Werkzeugaktivteile mit Innentemperierung zur Bearbeitung von Hochleistungsstählen" 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 Energy and Deutsches Zentrum für Luft- und Raumfahrt e. V. (German Aerospace Center) (DLR) as part of the program for promoting industrial cooperative research (IGF) based on a decision by the German Bundestag [8].

References

[1] Konradin-Verlag Robert Kohlhammer GmbH, Stanzstempel selbst nach 400 000 Hüben intakt, URL: https://industrieanzeiger.industrie.de/allgemein/stanzstempel-selbst-nach-400-000-hueben-intakt/, Accessed: 02.06.2026

[2] Asnafi, N., Rajalampi, J., Aspenberg, D., Alveflo, A. (2020): Production Tools Made by Additive Manufacturing Through Laser-based Powder Bed Fusion. BHM Berg- und Hüttenmännische Monatshefte 163(3), p. 125-136, doi: 10.1007/s00501-020-00961-8

[3] Bidare, P., Jiménez, A., Hassanin, H., Essa, K., (2022): Porosity, cracks, and mechanical properties of additively manufactured tooling alloys: a review. Advances in Manufacturing 10, p. 175-204, doi: 10.1007/s40436-021-00365-y

[4] Asnafi, N., (2021): Application of Laser-Based Powder Bed Fusion for Direct Metal Tooling. Metals 11(3), id. 458, doi: 10.3390/met11030458

[5] Chantzis, D., Liu, X., Politis, D. J., Fakir, O. E., Chua, T. Y., Shi, Z., Wang, L. (2020): Review on additive manufacturing of tooling for hot stamping. The International Journal of Advanced Manufacturing Technology 109(1-2), p. 87-107, doi: 10.1007/s00170-020-05622-1

[6] Das Fraunhofer-Institut für Werkzeugmaschinen und Umformtechnik IWU, Konturnahe Temperierung beim Presshärten, URL: https://www.iwu.fraunhofer.de/content/dam/iwu/de/documents/Infoblatt/Infoblatt-Konturnahe-Temperierung-beim-Pressh%C3%A4rten.pdf, Accessed: 02.06.2026

[7] Porstmann, S., Wannemacher, T., Drossel, W.-G., (2020): A comprehensive comparison of state-of-the-art manufacturing methods for fuel cell bipolar plates including anticipated future industry trends. Journal of Manufacturing Processes 60, p. 366–383, doi: 10.1016/j.jmapro.2020.10.041

[8] FGW Forschungsgemeinschaft Werkzeuge und Werkstoffe e. V., Additiv gefertigte Werkzeugaktivteile mit Innentemperierung zur Bearbeitung von Hochleistungsstählen, URL: https://www.fgw.de/wp-content/uploads/IGF_23013_N_AGAH_Additiv_gefertigte_Werkzeugteile-1.pdf, Accessed: 02.06.2026

Per-Dot Calibration Approach for Single-Shot 3D Sensors Using Structured Light

David Reese, Evgeny Degtyarev, Rico Nestler
Zentrum für Bild- und Signalverarbeitung e. V.
Werner-von-Siemens-Straße 12, 98693 Ilmenau
david.reese@zbs-ilmenau.de, evgeny.degtyarev@zbs-ilmenau.de, rico.nestler@zbs-ilmenau.de
www.zbs-ilmenau.de

Abstract: We present advances in the calibration pipeline for single-shot 3D sensors based on structured light dot pattern projection, building upon our previous work on projector-camera calibration via subpixel plane reconstruction. The primary contribution is a per-dot disparity-to-depth mapping that associates each projected dot with an individual hyperbolic depth model, enabling spatially resolved compensation of systematic depth errors across the full sensor working volume. To support the dense target sampling required for this calibration, we additionally present an improved chessboard target detection method related to OpenCV's Radon-transform-based corner detector, overcoming its key limitations regarding partially visible targets at image borders while also improving detection throughput for interactive use. Results are evaluated by comparing the systematic depth error of two sensor variants before and after the new calibration, providing a direct measure of improvement in 3D data quality.

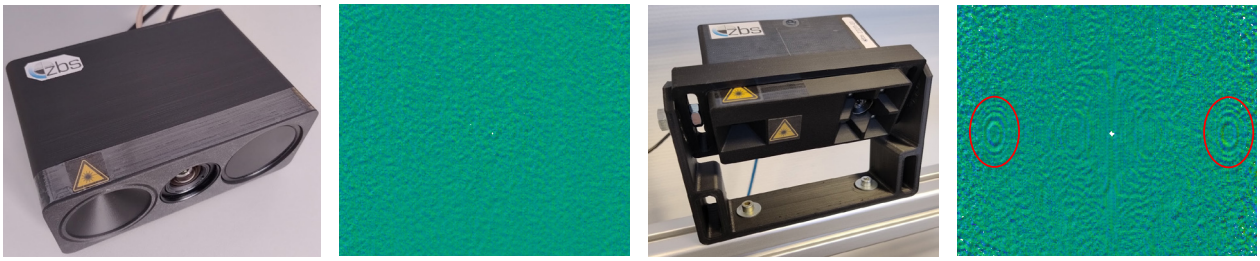
Keywords: Single-Shot 3D Imaging; Structured Light Calibration; Dot Pattern Projection; Per-Dot Disparity-to-Depth Mapping; Chessboard Target Detection; Radon Target Detection

1 Scientific Background and Motivation

Single-shot 3D sensors based on structured light dot pattern projection offer practical advantages for compact, real-time depth sensing. Since no sequential projection is required, these systems are inherently robust to scene motion. To exploit their full potential, for instance in 3D scanning and quality inspection applications, accurate geometric calibration is essential, both of the projector-camera system [1] and of the sensor's depth response across its working volume. This article evaluates two sensor variants built around laser projectors with diffractive optical elements (DOEs) projecting a fixed pseudorandom dot pattern. Both systems were intrinsically and extrinsically calibrated following the procedure described in [1] and using our calibration toolkit 3D-EasyCalib™ [2].

System1 (Fig. 1a) uses a single 830 nm dot pattern projector together with a NIR camera and additionally incorporates an RGB camera in the center. Instead of an RGB camera, *System2* (Fig. 1c) integrates a flood illumination unit alongside a 940 nm dot pattern projector, acquiring dot-pattern and flood-illuminated frames in alternation during operation. System2 is built with tighter size and cost constraints in mind, using a smaller back-illuminated sensor and a compact M12 lens. The two systems are furthermore configured for different working-distance ranges, System1's being shifted somewhat farther than System2's.

Figs. 1b and 1d show, for a flat reference surface, the color-coded directions of the reconstructed surface normal vectors when imaged with pattern I (System1) and pattern II (System2). Deviations from a uniform color indicate depth error, which is comparatively small for pattern I, whereas pattern II shows markedly larger deviations with a distinct wave-like structure. The generally higher error level of System2 is consistent with its lower-cost camera components. The wave-like structure, however, could clearly be attributed to the projector, being stationary in its view and independent of distance. System2's projector also covers a wider field of illumination, with a diagonal extent of approximately $86°$ versus $78°$. We suspect this contributes to the effect: a periodic DOE can only direct light into diffraction orders that are equally spaced in direction-cosine space [3], so that in IFTA-based design each dot of the target pattern is placed at the nearest available order [4]. The resulting residual position errors vary systematically across the field and form a moiré-like structure, which a wider field of illumination is likely to amplify.



(a) (b) (c) (d)

Fig. 1: Sensor heads of System1 (a), with an additional RGB camera in the center, and System2 (c), with an additional flood illumination unit. (b) and (d) show surface normal images (with color-coded normal vector directions), reconstructed from the projector reference frames of the respective systems. In (d), regions where the mentioned wave-like structures are best visible are marked in red. Gaps in the otherwise visible structure of the projected pattern are inpainted for better visualization.

Across both systems, 3D point cloud generation begins by locating, for each projected dot, its corresponding position in the NIR image. The displacement between a dot's position in the projected pattern and its observed position in the image yields that dot's disparity. Its component along the sensor rows varies with the distance of the illuminated surface and is the quantity used for the disparity-to-depth mapping. The identified point is then reprojected into space either via triangulation, using the shared intrinsic and extrinsic calibration obtained as in [1], or via the per-dot disparity-to-depth mapping presented in this article.

In our previous work [1], we demonstrated that replacing chessboard corner-based projector correspondences with directly localized dot centers yields more accurate projector-camera calibration with improved spatial consistency of the reconstructed 3D geometry. However, the standard parametric camera model, which describes sensor behaviour globally within the measurement volume using a limited set of intrinsic and extrinsic parameters, cannot capture position-dependent deviations in the disparity-to-depth conversion for individual projected dots, arising for example from imperfect lens geometries, manufacturing tolerances in the projector optics, or the DOE-related effects discussed above. Per-dot calibration addresses this remaining source of systematic depth error by modeling the disparity-to-depth relationship individually for each dot of the projected pattern.

To estimate the per-dot model parameters, target observations at many distances are required, covering the full sensor working volume, including boundary regions where the target is inevitably clipped by the image frame. Standard chessboard detectors, including OpenCV's `findChessboardCornersSB` as well as ChArUco-based methods, systematically discard valid border corners: the former requires detected corners to form a complete rectangular grid, and the latter only reports corners when both adjacent ArUco markers are fully visible. This motivated the improved target detection method described in Section 2.

2 Methods Used

Per-dot disparity-to-depth calibration

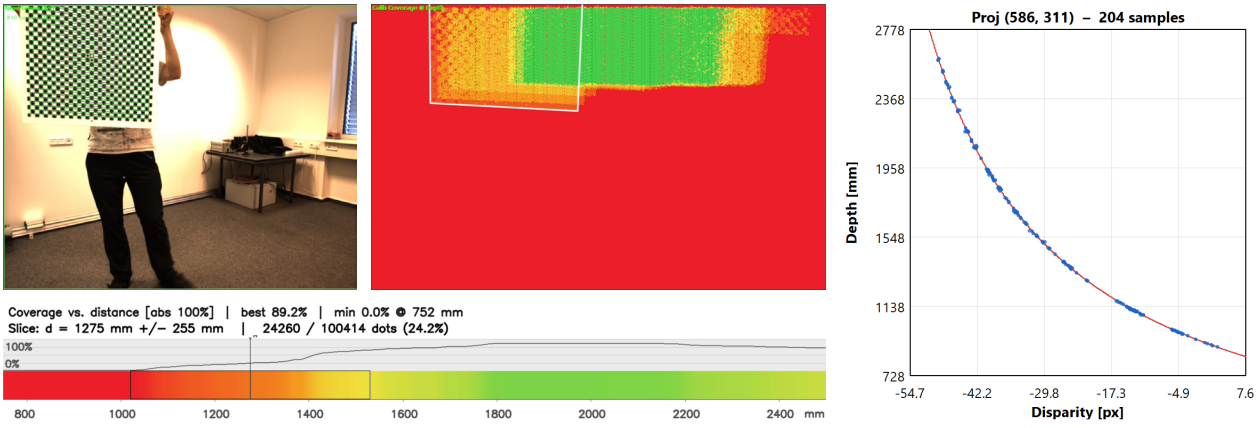
For a simplified triangulation-based system with parallel optical axes and baseline B, treating the projector-camera pair analogously to a stereo camera system and assuming a pinhole camera model in the sense of Tsai [5], depth Z and disparity d are related by

$$Z(d) = \frac{fB}{d},$$

with f denoting the camera constant, also referred to as the principal distance. This idealized relationship is a rectangular hyperbola in d (i.e. $Zd = fB = \text{const}$), motivating our choice of a hyperbolic function as the underlying model for the disparity-to-depth mapping. Real systems, however, deviate from this idealized geometry, e.g. through non-parallel optical axes, lens distortion, or other projector-specific systematic effects. A single hyperbola parameterized only by f and B can therefore not capture the true, spatially varying relationship. We therefore model the disparity-to-depth relationship for each projected dot by an individually fitted hyperbolic function, retaining the functional

form motivated above while allowing its parameters to vary across the pattern and thereby absorb these local systematic deviations. Parameters are estimated by capturing a planar target at multiple known distances throughout the working volume and fitting the model per dot using the reconstructed depth values. Data collection is supported by a semi-automatic sampling procedure: while the target is moved manually, frame acquisition is automated based on criteria such as target orientation and motion velocity, with visualization of spatial coverage.

The depth values to be sampled are defined in the NIR camera's coordinate system. For each frame, the target's pose is estimated from its detected corners and the intrinsic camera parameters using OpenCV's `solvePnP`. If detection succeeds and the pose satisfies the acceptance criteria discussed below, it defines a reference plane in the NIR camera frame, and intersecting each projector ray with it gives the corresponding surface point, whose distance along the NIR camera's optical axis serves as the reference depth for that dot. Obtaining the target pose reliably, and expressed in this frame, differs between the two sensor variants and constitutes the main practical difficulty of the sampling procedure. System1's auxiliary RGB camera with NIR-cutoff filter allows the target to be detected independently of the dot pattern, with its pose transformed into the NIR camera frame via the extrinsic RGB-to-NIR calibration, so that pose and depth data originate from the same instant. System2 has no such auxiliary view. Target detection instead runs on flood-illuminated frames acquired in alternation with the dot-pattern frames used for depth reconstruction, introducing a temporal offset between the two. Since any target motion during this offset directly biases the reconstructed reference depth, the sampling procedure discourages fast motion in particular along the sensor's depth axis, as an example of such a mitigating constraint. A lateral displacement of comparable magnitude affects the reference depth far less. Frames with excessive target tilt relative to the optical axis are excluded as well, since oblique views increase the sensitivity of both the plane fit and the resulting ray-plane intersection to small corner-localization errors. Fig. 2a shows the live visualization presented to the operator during sampling: detected target corners in the current camera frame (top left), together with two complementary views of accumulated sampling coverage, namely a projector-view slice restricted to a tolerance band around the current target distance (top right) and a profile of coverage against distance across the full working volume (bottom). This lets under-sampled regions of the pattern or working range be identified and revisited within the same session. Fig. 2b shows, alongside this, the accumulated disparity-depth samples and fitted hyperbolic model for a single, interactively selected dot, supporting a live check of individual per-dot fit quality.



(a) (b)

Fig. 2: Live sampling visualization. (a) Operator view during sampling, combining detected target corners in the current camera frame (top left), sampling coverage in projector view restricted to a tolerance band around the target's current distance (top right, color-coded from not yet sampled (red) to fully sampled (green)), and the corresponding coverage profile against distance across the working volume (bottom). (b) Disparity-depth samples and fitted hyperbolic model for a single, manually selected dot after the completed sampling process.

Reliable operator feedback during sampling depends on target detection that is both fast enough for interactive frame rates and tolerant of the frequent border-clipping that occurs while sweeping the target through the working volume, particularly near the edges of the pattern. Standard chessboard detectors satisfy neither requirement well, as discussed next.

Improved chessboard target detection

Our detection method uses the same corner feature extraction as OpenCV's `findChessboardCornersSB` and is related to it in its overall structure: a fast, point-symmetric candidate corner detector evaluating a localized Radon transform along multiple orientations (`FastX` in OpenCV's implementation) provides candidates for sub-pixel corner locations, following [6], whose evaluation of corner localization accuracy therefore applies comparably here. The primary implementation changes are a generalization of the output data structure and the underlying detection pipeline: rather than requiring detected corners to form a complete rectangular grid, our detector returns all corners whose board coordinates can be unambiguously determined via a set of circular markers, including those at clipped board edges and under partial occlusion elsewhere on the target. The `FastX` computation is additionally ported to OpenCL for GPU acceleration, improving detection throughput for interactive calibration workflows.

OpenCV's `findChessboardCornersSB` (tested with OpenCV 4.11.0 in C++) does in principle support recovering corners from partially visible targets with circular markers when using the `CALIB_CB_MARKER` and `CALIB_CB_LARGER` flags together with setting a small target size (e.g. 3×3 corners) as parameters, with the actual target being larger. The `meta` output array reports, per corner, whether it coincides with a marker, which resolves the board origin. This combination is comparatively undocumented and, in our experience, not widely known. However, the corners returned this way always form a single, complete rectangular sub-grid of the physical board: if any corner within that sub-grid cannot be localized, the corresponding row and column, and everything beyond it, is discarded as well, even where those corners are themselves unaffected and readily detectable. Figs. 3a and 3c show this for a border-clipped and a glare-affected target capture, respectively.



(a) (b) (c) (d)

Fig. 3: Comparison between the OpenCV detector `findChessboardCornersSB` (a, c) and the proposed method (b, d) for two failure cases: a target partially clipped by the image border (a, b) and a target with specular glare, while the dot markers remain visible (c, d). Detected corners are color-coded by their deviation from the respective estimated grid model, from green (small deviation) to red (large deviation).

Our detector shares its initial stages with `findChessboardCornersSB`: `FastX` supplies sub-pixel corner candidates across multiple scales, and initial board candidates are generated from these in essentially the same way. The differences lie in how such a candidate is subsequently grown into a board. Growth proceeds outward from a small seed region: for each board position adjacent to the already-detected set, an expected image location is predicted from previously detected neighbours along each of the two board axes. Where predictions from both axes are available, they must agree before the corner is accepted. Predicted locations are matched against the `FastX` candidates. Because growth follows connectivity rather than filling a bounding rectangle, the detected region can take on arbitrary, non-rectangular shapes instead of collapsing to the largest fully intact sub-grid.

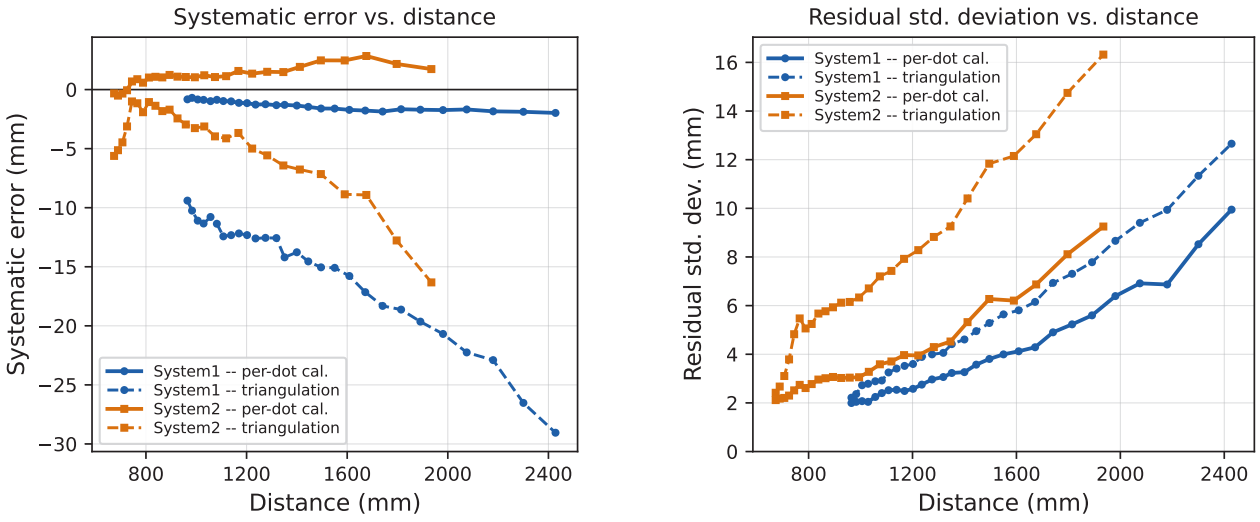
Where growth stalls at gaps, e.g. caused by occlusion, glare or the marker cutout itself, a second pass interpolates the missing positions from the surrounding geometry, again requiring agreement between independent estimates so that the fill remains confined to regions with supporting context. Interpolated corners are flagged as such in the output, so that downstream steps needing precise, image-measured locations – such as the per-dot calibration presented in this paper – can rely solely on directly detected corners, while interpolated ones still support board connectivity and origin resolution. Together, these measures recover the corners discarded by the original detector in both failure cases, as shown in Figs. 3b and 3d.

A related, more general approach to recovering occluded checkerboard corners is presented by Hillen et al. [7],

who fit a Gaussian process regressing image pixel coordinates from board coordinates as a detector-agnostic post-processing step on top of an initial corner detection, using the fitted model to fill in occluded or missing corners, extend the board beyond the image border, and re-assign initially misassigned corners. In contrast, our propagation is not a separate post-processing stage decoupled from the underlying detector, but performs interpolation directly during growth, using local, per-direction geometric estimates rather than a globally fitted regression model. Combined with the spatial keypoint index, this keeps the computational cost low enough for our targeted use case of interactive, high-throughput disparity-to-depth sampling.

3 Evaluation and Results

Depth accuracy of the 3D sensor systems is assessed using the same target-as-reference-plane principle as the sampling procedure (see Methods), but for measurement rather than calibration: for each accepted frame, the depth predicted from every matched dot correspondence is compared against the reference depth given by the intersection of the corresponding projector ray with the detected target plane, yielding a per-point residual. Two depth estimation methods are evaluated side by side on the same correspondences: the per-dot calibration presented in this paper, and the underlying triangulation-based approach it replaces, i.e. depth reconstruction from the shared intrinsic and extrinsic calibration [1] without per-dot disparity-to-depth mapping. Residuals are aggregated over many frames into the mean residual (systematic error) and the residual standard deviation at each distance, evaluated separately for each of the two systems described above. The target was kept within the central field of view so that each distance is evaluated over a comparable set of dots, since the aim here is to demonstrate the principle on functional prototype systems rather than to characterize them exhaustively. Owing to the different working-distance ranges noted above, the evaluated distances differ between the two systems. Fig. 4 shows both quantities across the full evaluated range, for both systems and both methods.



(a) (b)

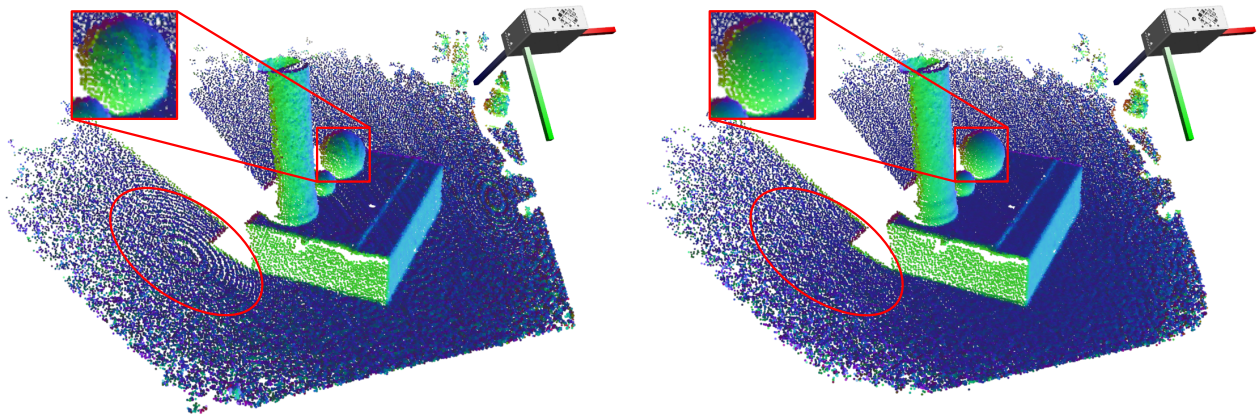
Fig. 4: Systematic error (a) and residual standard deviation (b) versus distance, for both systems and both depth estimation methods, over the full evaluated range. Solid lines: per-dot calibration. Dashed lines: triangulation. As described above, System1 and System2 differ in enough other respects that their curves should not be read as a direct ranking between the two designs. The relevant comparison is calibrated versus uncalibrated within each system.

For both systems, the per-dot calibration reduces the systematic error by roughly an order of magnitude at the far end of the working range. The residual standard deviation is likewise clearly reduced, by roughly a factor of two across most of System2's working range and by roughly a factor of 1.3 to 1.4 for System1. Residual error which is partially caused by sensor noise grows with distance for either method, as is expected for a triangulation-based system. Systematic error is largely reduced by the per-dot calibration, especially for larger distances.

4 Discussion

At each distance, residuals are pooled over all evaluated dots before computing their mean and standard deviation. The mean therefore reflects error components common to all dots, such as a global depth offset. Deviations that vary across the field of view, such as the wave-like structure, instead show up in the standard deviation. This structure is systematic: it is fixed in the projector's view and does not change with distance, so an individually fitted model per dot can absorb it. The marked reduction in standard deviation is therefore mainly due to the correction of this structure, and is correspondingly larger for System2, whose distortion was more pronounced. What remains is a stochastic component, caused for example by sensor noise. It grows with distance, since the disparity-to-depth mapping becomes increasingly sensitive to disparity errors. Being random, it cannot be compensated by any calibration.

Although the quantitative evaluation is restricted to the central region of the camera's field of view, the benefit of the per-dot calibration is not confined to that region. Fig. 5 shows pointclouds with color-coded surface normal directions for System2 on a scene containing several non-planar objects, reconstructed once by triangulation and once using the per-dot calibration. The wave-like structure characteristic of the uncalibrated system is visibly reduced across the whole field of view, including areas well away from the image centre, and object surfaces appear correspondingly more uniform.



(a) (b)

Fig. 5: Pointclouds with color-coded surface normal vector directions for System2 on a scene containing several non-planar objects, reconstructed via triangulation (a) and via the per-dot disparity-to-depth calibration (b). Notable differences are highlighted in red: The wave-like structure visible in (a) practically disappears with the per-dot disparity-to-depth calibration. Also note the improved surface quality e.g. on the larger sphere.

Strictly speaking, the per-dot calibration does not depend on the full calibration procedure of [1]: only the NIR camera's intrinsic parameters are required, to back-project each identified dot into the NIR camera's coordinate system using its per-dot depth and its direction from the intrinsic parameters. The extrinsic parameters remain necessary, however, whenever the point cloud is required in the projector's coordinate system, where the fixed pattern structure is directly available and supports efficient downstream processing such as outlier removal and surface normal computation.

The per-dot calibration involves a substantially larger overall parameter set than the handful of global intrinsic and extrinsic parameters used by triangulation. The per-point runtime cost can nevertheless be lower: in our case it replaces an iterative per-point undistortion and triangulation which takes considerably longer than a simple hyperbolic parameter lookup and evaluation. Its accuracy is nonetheless bounded by the quality of the underlying camera calibration, since the reference pose used during sampling is derived from it: errors in the intrinsic parameters of both systems, and additionally in System1's extrinsic RGB-to-NIR calibration, propagate into that pose. Real lenses moreover deviate somewhat from an ideal pinhole model, so distance-dependent intrinsic calibrations could be a worthwhile refinement where very high accuracy is required.

It is worth emphasising that the improvements reported here were obtained with a purely manual sampling procedure. Despite the error contributions this entails, a substantial reduction of systematic depth error is achieved,

which suggests that automating the acquisition offers room for further gains, for instance by moving the camera on a pan-tilt head and linear rail rather than the target.

5 Conclusions

We presented a per-dot disparity-to-depth calibration for single-shot structured-light 3D sensors, together with an improved chessboard target detection method that supports the dense, boundary-inclusive target sampling this calibration requires. Evaluated on two sensor variants, the approach substantially reduces systematic depth error compared to the underlying triangulation-based reconstruction, particularly at longer working distances, while a semi-automatic sampling procedure keeps the additional calibration effort manageable. These gains directly benefit applications such as 3D scanning and quality inspection, where spatially accurate depth across the full working volume is essential.

The individual contributions are not tied to the two sensor variants evaluated here. The per-dot calibration requires only a fixed, reproducible projected pattern and a way to establish correspondences, and is therefore applicable to such sensors in general; it is particularly attractive wherever cost constraints on the optical components lead to systematic deviations that global parametric models cannot capture. The semi-automatic sampling procedure and its coverage visualization transfer to any calibration task requiring dense, spatially resolved observations across a working volume. The improved target detection, finally, is of general use for camera calibration and extends the range of methods available in our calibration toolkit 3D-EasyCalib™ [2].

The results presented in this article are based on our work within the joint project "Implicit mobile human-robot communication for spatial action coordination with action-specific semantic environment modelling" (EDISON), grant no. 13N17577. The project is funded by the German Federal Ministry of Research, Technology and Space (BMFTR). EDISON takes place within the framework of the funding program "Photonik für die digitalisierte und automatisierte Produktion".

References

[1] D. Reese, E. Degtyarev, and R. Nestler, "Accurate projector-camera-calibration from structured light dot patterns via subpixel plane reconstruction," in *Anwendungsbezogener Workshop „3D in Science & Applications" (3D-iSA), Tagungsband, GFaI e. V.*, (Berlin), 2025.

[2] D. Vehar, R. Nestler, and K.-H. Franke, "3D-EasyCalib – Toolkit zur geometrischen Kalibrierung von Kameras und Robotern," in *22. Anwendungsbezogener Workshop zur Erfassung, Modellierung, Verarbeitung und Auswertung von 3D-Daten „3D-NordOst"*, (Berlin), p. 15, GFaI Gesellschaft zur Förderung angewandter Informatik e. V., Dec. 2019.

[3] J. E. Harvey and R. N. Pfisterer, "Understanding diffraction grating behavior: including conical diffraction and rayleigh anomalies from transmission gratings," *Optical Engineering*, vol. 58, no. 8, p. 087105, 2019.

[4] X. Liu, Y. Hu, S. Tu, C. Kuang, X. Liu, and X. Hao, "Fast generation of arbitrary optical focus array," *Optics and Lasers in Engineering*, vol. 162, p. 107405, 2023.

[5] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal on Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.

[6] A. Duda and U. Frese, "Accurate detection and localization of checkerboard corners for calibration," in *British Machine Vision Conference (BMVC)*, (Newcastle, UK), p. 126, 2018.

[7] M. Hillen, I. D. Boi, T. D. Kerf, S. Sels, E. C. D. L. Hoz, J. Gladines, G. Steenackers, R. Penne, and S. Vanlanduit, "Enhanced checkerboard detection using gaussian processes," *Mathematics*, vol. 11, no. 22, p. 4568, 2023.

Markerbasiertes Trackingsystem zur roboterunabhängigen Trajektoriengenerierung für automatisierte Beschichtungsprozesse

Wolff, Martin[1], Hohnhäuser, Benjamin[1], Seifert, Lukas[2]

[1]GFaI e.V., Volmerstraße 3, 12489 Berlin, wolff@gfai.de, hohnhaeuser@gfai.de, www.gfai.de

[2]THÜMA Maschinenbau und Service GmbH, Pfarrstraße 8, 99867 Gotha, l.seifert@thuema.de

Abstract: In vielen Bereichen der industriellen Fertigung nimmt der Automatisierungsgrad stetig zu, wobei der Grad der Automation oft mit abnehmender Losgröße sinkt. Insbesondere bei kleinen Losgrößen ist eine Automatisierung von Beschichtungsprozessen aus wirtschaftlicher Sicht häufig schwer umsetzbar, obwohl hier das Potenzial für Einsparungen in Bezug auf Material, Zeit und Energie besonders hoch ist. Im Rahmen des FuE-Projekts AuLaTeach wurde ein System entwickelt, das die manuelle Bewegung der Sprühpistole bei der Erzeugung von Sprühbahnen in Echtzeit erfasst und wichtige Prozessparameter als roboterunabhängige Trajektorie aufzeichnet. Durch eine Kombination aus optischen Markern und einer inertialen Messeinheit (IMU) kann durch den Einsatz eines Multikamerasystems die zeitabhängige Pose der Sprühpistole präzise bestimmt werden.

Keywords: Markerverfolgung; Multikamerasystem; Trajektoriengenerierung; Beschichtungsautomatisierung; Posenschätzung; IMU-Fusion; Kalibrierwürfel; Sprühbahnerfassung

1 Thema

In vielen Bereichen der industriellen Fertigung nimmt der Automatisierungsgrad stetig zu, wobei der Grad der Automation oft mit abnehmender Losgröße sinkt. Insbesondere bei kleinen Losgrößen ist eine Automatisierung von Beschichtungsprozessen aus wirtschaftlicher Sicht häufig schwer umsetzbar, obwohl hier das Potenzial für Einsparungen in Bezug auf Material, Zeit und Energie besonders hoch ist. Dieser Beitrag stellt das Forschungs- und Entwicklungsprojekt AuLaTeach vor, das im Rahmen des ZIM-Innovationsnetzwerks AI4Tech in Zusammenarbeit zwischen der GFaI im Bereich Bildverarbeitung und industrielle Anwendung (BVIA) und der THÜMA Maschinenbau und Service GmbH durchgeführt wurde. Das System erfasst manuelle Lackiervorgänge vollständig und überführt diese automatisch in ausführbare Roboterprogramme. Ein erfahrener Lackierer führt den Beschichtungsvorgang dabei einmalig von Hand aus, wobei das System Bewegungsbahn und Prozessparameter aufzeichnet und daraus eine roboterunabhängige Trajektorie erzeugt, die anschließend flexibel an beliebige Industrieroboter angebunden werden kann.

2 Wissenschaftlicher Hintergrund und Motivation

Das manuelle Spritzlackieren lässt sich in variantenreichen Fertigungsumgebungen mit kleinen Losgrößen nur schwer automatisieren. Klassische Programmiermethoden wie Offline-Programmierung oder manuelles Teach-In erfordern erheblichen Einrichtungs- und Programmieraufwand und sind ungeeignet, wenn sich Werkstückgeometrien häufig ändern oder die Stückzahlen gering sind. Gleichzeitig verfügen Lackierfachkräfte über implizites Prozesswissen, das bislang nicht strukturiert und reproduzierbar auf Robotersysteme übertragen werden kann. AuLaTeach schließt diese Lücke durch die Entwicklung einer Methodik und Hardware-Software-Infrastruktur, die das Erfahrungswissen der Lackierfachkraft erfasst, digitalisiert und in ein roboterausführbares Format überführt, einschließlich vollständiger Bewegungstrajektorien sowie prozessrelevanter Parameter wie Luftdruck, Materialfluss und Sprühzustand.

Für die Erfassung handgeführter Werkzeugbewegungen kommen grundsätzlich verschiedene Trackingansätze in Betracht. Kommerzielle markerbasierte Motion-Capture-Systeme erreichen Positionsgenauigkeiten im Submillimeterbereich unter statischen und unter 2 mm unter dynamischen Bedingungen [1], sind jedoch mit hohen Investitionskosten verbunden und erfassen ausschließlich die Bewegung, nicht aber die zugehörigen Prozessparameter. Kostengünstige Trackingsysteme aus dem VR-Bereich wurden bereits für die Programmierung von Lackierrobotern

durch Vormachen eingesetzt [2]. Rein optische Verfahren ohne Inertialsensorik verlieren bei Verdeckungen die Pose, während rein inertiale Verfahren einer Integrationsdrift unterliegen. Der hier vorgestellte Ansatz kombiniert ein für die Lackierumgebung ausgelegtes Multikamerasystem mit einer IMU und integriert die Prozessdatenerfassung unmittelbar in die Trajektorie.

3 Verwendete Methoden

3.1 Kamerasystem und Kalibrierung

Kernkomponente des Systems ist ein markerbasiertes Multikamera-Trackingsystem bestehend aus acht Industriekameras (1280 × 960 Pixel, Bildrate ca. 5,6 fps), die in einer vollständig abgeschlossenen Lackierkabine fest montiert sind. Die Kameras erfassen ein an der Sprühpistole montiertes LED-Markersystem und ermöglichen so eine kontinuierliche räumliche Verfolgung der Pistolenbewegung. Die Kalibrierung des Systems erfolgt in zwei aufeinanderfolgenden Stufen. Zunächst wird jede Kamera einzeln intrinsisch nach der Zhang-Methode kalibriert, wobei die inneren optischen Parameter wie Brennweite, Hauptpunkt und Verzeichnungskoeffizienten bestimmt werden [3].

Die extrinsische Einmessung des Gesamtsystems erfolgt mithilfe eines eigens entwickelten 3D-Kalibrierkörpers: einem Würfel, auf dessen sechs Flächen je ein Diamond-ChArUco-Board aufgebracht ist [4, 5]. Für jede Kamera werden Translations- und Rotationsvektoren relativ zum Kalibrierkörper bestimmt. Anschließend wird der Levenberg-Marquardt-Algorithmus zur nichtlinearen Optimierung der Transformationen aller acht Kameras in ein gemeinsames globales Koordinatensystem eingesetzt. Das Ergebnis ist ein vollständig eingemessenes Mehrkamerasystem, das eine präzise räumliche Rekonstruktion der Lackierbahn ermöglicht.

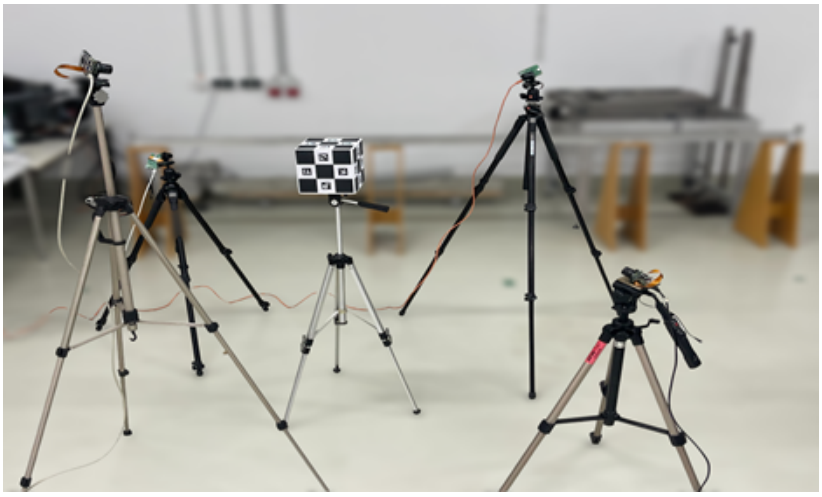


Abb. 1: Aufbau des Multikamerasystems mit 3D-Kalibrierwürfel (Diamond-ChArUco-Boards) zur extrinsischen Systemkalibrierung.

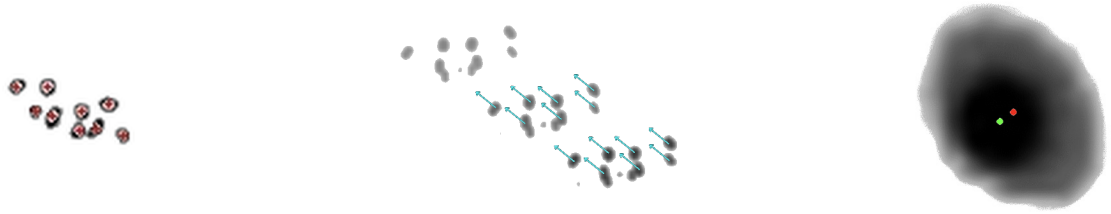
3.2 Optisches Tracking, Posenschätzung und Target-Design

Das LED-basierte Marker-Tracking verfolgt kontinuierlich die Position und Orientierung der Lackierpistole. Jeder Kamera ist ein optischer Bandpassfilter vorgelagert, der selektiv nur das Emissionsspektrum der aktiven LED-Marker (625 nm) durchlässt und Störlicht aus dem Lackierumfeld effektiv unterdrückt. Die Anordnung der aktiven Marker auf dem Target erfordert einen sorgfältig abgewogenen Kompromiss zwischen Erkennungsqualität und Handhabbarkeit für den Lackierer. Dazu wurde eine Halterung entwickelt, die es ermöglicht, verschiedene Markerpositionen und -muster flexibel zu erproben. Die Evaluation zeigte, dass verschiedenfarbige LEDs die Genauigkeit der Posenschätzung nicht verbessern und zudem den Einsatz schmalbandiger Filter verhindern. Daher werden acht monochrome rote LEDs eingesetzt.

Die Bildaufnahme erfolgt über je eine eigene Recorder-Instanz pro Kamera. Da eine hardwareseitige Synchronisation aller Kameras die erreichbare Bildrate reduzieren würde, arbeiten die Kameras im Freilauf. Jedes Bild erhält

unmittelbar einen Zeitstempel, und die Markerpositionen zwischen den Aufnahmezeitpunkten werden durch lineare Interpolation im Bild bestimmt. Im laufenden Betrieb werden statt der Rohbilder nur die detektierten Marker mit Schwerpunkt und Intensität gespeichert, was Datenmenge und Rechenlast deutlich reduziert.

Die Marker werden im Graustufenbild mittels Flood-Fill-Algorithmus segmentiert [6]. Eine Binarisierung wird bewusst vermieden, da gleiche Pixelgewichtungen zu fehlerhaften Schwerpunktsberechnungen führen. Stattdessen wird der Schwerpunkt jedes Markers durch einen intensitätsgewichteten Mittelwert subpixelgenau bestimmt. Die Verfolgung der Marker zwischen aufeinanderfolgenden Frames erfolgt zusätzlich mithilfe von optischem Fluss nach Lucas-Kanade, wodurch auch bei kurzzeitig fehlenden Markerpositionen eine kontinuierliche Verfolgung der Bewegungsbahn gewährleistet wird. Zu Beginn jeder Aufnahme wird die Pistole in eine definierte Ausgangsposition gebracht, aus der die initiale Pose eindeutig bestimmt wird. Die finale 6D-Pose des Targets ergibt sich aus einer Perspective-n-Point-Lösung (PnP) [7], die bekannte 3D-Referenzpositionen der LEDs mit beobachteten 2D-Positionen in den Kamerabildern in Übereinstimmung bringt. Für eine eindeutige Lösung sind mindestens drei sichtbare Marker erforderlich, wobei der Einsatz von acht Kameras die Wahrscheinlichkeit ausreichender Markersichtbarkeit auch bei teilweisen Verdeckungen maximiert. Mehrdeutigkeiten bei der Zuordnung der Marker werden über die bekannte LED-Geometrie des Targets sowie eine Plausibilitätsprüfung anhand der Trajektorienhistorie aufgelöst. Jede Pose erhält zudem eine Qualitätsbewertung aus Anzahl erkannter Marker und Plausibilität gegenüber der vorherigen Pose. Die Markerdaten der im Freilauf arbeitenden Kameras werden durch Interpolation auf einen gemeinsamen 10-Hz-Takt gebracht, mit dem die kamerabasierte 6D-Pose bestimmt wird.



(a) (b) (c)

Abb. 2: Markerdetektion (invertierte Darstellung): (a) LED-Erkennung nach optischer Filterung, (b) Markerverfolgung mit Bewegungsvektoren, (c) Unterschied zwischen binarisierungsbasiertem Schwerpunkt (rot) und intensitätsgewichtetem Mittelwert (grün).

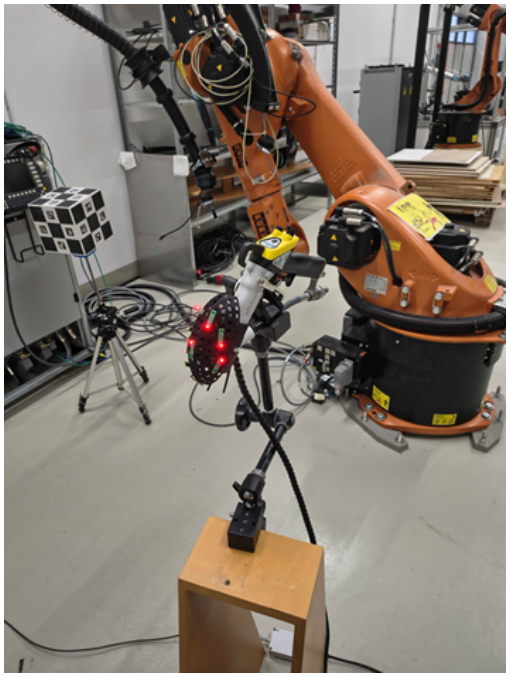


Abb. 3: Sprühpistole mit LED-Markersystem im Demonstratoraufbau, im Hintergrund der Kalibrierwürfel.

3.3 IMU-Anbindung und Datenfusion

Ergänzend zum kamerabasierten Tracking wurde eine IMU direkt an der Lackierpistole befestigt, die kontinuierlich translatorische Beschleunigungen sowie Drehraten in allen drei Raumachsen mit bis zu 100 Hz erfasst. Eine ausschließlich IMU-basierte Positionsverfolgung ist nicht praktikabel, da die durch doppelte Integration der Beschleunigungsdaten berechnete Position einer kontinuierlich anwachsenden Integrationsdrift unterliegt. Die IMU dient daher als ergänzende Datenquelle zur Überbrückung von Ausfällen des kamerabasierten Trackings von typischerweise bis zu zwei Sekunden. Unterschreitet die Qualitätsbewertung der Kamerapose einen Schwellwert, etwa bei zu wenigen sichtbaren Markern oder unplausiblen Posen, übernimmt die IMU kurzzeitig die Posenschätzung. Ein Kalman-Filter [8] in der Implementierung von OpenCV fusioniert beide Datenquellen, indem die hochfrequenten IMU-Daten zur Vorhersage der nächsten Pose genutzt und anschließend durch eintreffende Kameradaten korrigiert werden. Die dynamische Gewichtung beider Quellen verhindert dabei Sprünge in der aufgezeichneten Trajektorie.

3.4 Prozessparameter und Trajektoriengenerierung

Parallel zum Tracking werden alle prozessseitigen Größen synchron erfasst: Drücke, Durchflussmengen, Ventilstellungen sowie der Sprühzustand der Lackierpistole werden mit 10 Hz aufgezeichnet. Jedem Datensatz wird ein eindeutiger Zeitstempel zugeordnet, der eine zeitgenaue Synchronisation und Interpolation zwischen den verschiedenen Datenquellen ermöglicht. Aus der aufgezeichneten Trajektorie werden Stützpunkte in einem konfigurierbaren Raster extrahiert, wahlweise zeitbasiert (typisch 100 ms) oder nach räumlichen Abstandskriterien. Zu jedem Stützpunkt werden die 6D-Pose sowie alle interpolierten Prozessparameter als Attribute hinterlegt. Die resultierende Bahn wird als strikte Werkzeug-Werkstück-Relation gespeichert, die die Relativbewegung zwischen Sprühwerkzeug und Werkstück unabhängig von der jeweiligen Aufnahmeposition beschreibt und damit direkt auf beliebige andere Roboterkinematiken übertragbar ist.

3.5 Roboterprogrammgenerierung und grafische Benutzeroberfläche

Eine automatisierte Nachverarbeitungspipeline überführt die Werkzeug-Werkstück-Relation in ein ausführbares Roboterprogramm für die jeweilige Zielkinematik. Dazu wird die Bahn zunächst anhand von Richtungs- und Orientierungsänderungen sowie Wechseln des Abzugsstatus in Segmente zerlegt, wobei Segmente ohne Sprühvorgang als Leerfahrten gesondert behandelt werden. Anschließend werden mehrere Validierungsschritte durchlaufen: Erreichbarkeit mittels inverser Kinematik, Gelenkgrenzeneinhaltung, Singularitätsnähe sowie potenzielle Kollisionen mit der Umgebungsgeometrie werden geprüft und kritische Bahnabschnitte dem Anwender visualisiert. Gültige Segmente werden über automatisch erzeugte, kollisionsfreie Verbindungsbahnen zusammengesetzt, auf ein festes Zeitraster (z. B. 40 ms) umgetastet und im roboterspezifischen Format exportiert. Eine grafische Benutzeroberfläche ermöglicht die interaktive 3D-Inspektion und manuelle Korrektur der Trajektorien. Kernfunktion ist eine integrierte interaktive Schichtdickensimulation, die das zu erwartende Beschichtungsergebnis auf Basis eines 3D-Werkstückmodells berechnet und nach jeder Bahnänderung aktualisiert, sodass eine simulationsgestützte Optimierung vor dem realen Robotereinsatz möglich ist.

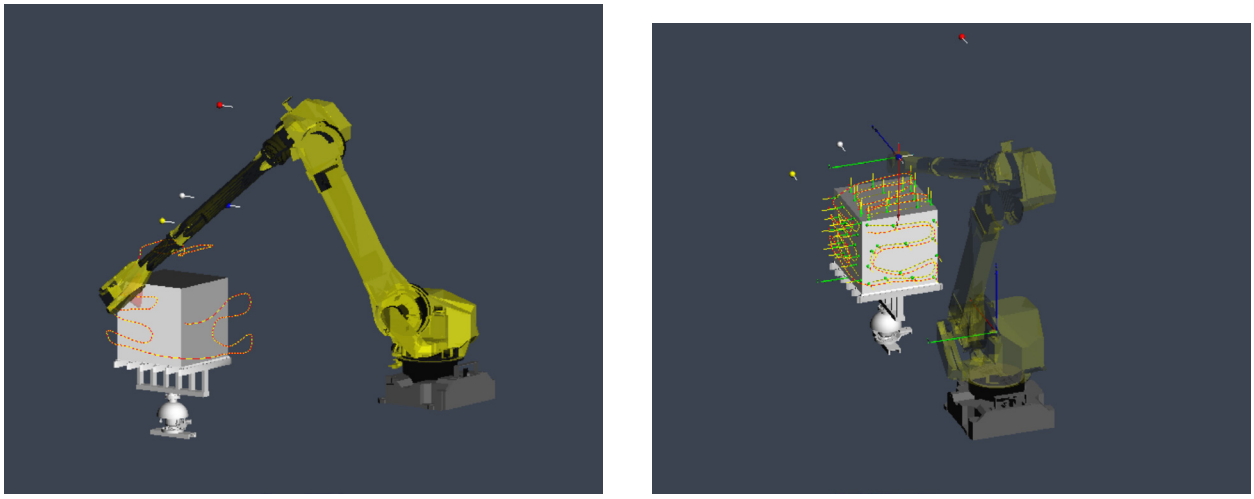


Abb. 4: GUI-Ansichten der Offline-Simulationsumgebung: Trajektorie mit Lackierbahn am Werkstück (links) und Detailansicht mit Orientierungsvektoren (rechts).

4 Ergebnisse

Das entwickelte Trackingsystem ermöglicht die präzise Erfassung der zeitabhängigen Pose der Sprühpistole im Raum und erzeugt daraus eine vollständige, roboter- und anlagenunabhängige Lackiertrajektorie. Die gemessene Positionsgenauigkeit von 3,8 mm und die Orientierungsgenauigkeit von 0,7° liegen innerhalb der definierten Zielgrößen von < 5 mm bzw. $< 1°$. Zur Bestimmung der Trackinggenauigkeit wurde das Target an einem KUKA-Industrieroboter montiert (vgl. Abb. 3) und entlang definierter linearer Fahrten in X-, Y- und Z-Richtung sowie einer Kurvenfahrt bewegt. Die Fahrten wurden mehrfach und mit unterschiedlichen Kamerapositionen wiederholt. Als Referenz dienten die über die Robot-Sensor-Interface-Schnittstelle (RSI) übertragenen Roboterposen, die mit den vom Trackingsystem berechneten Posen verglichen wurden [6]. Die generierte Trajektorie weist eine räumliche Auflösung von unter 1 mm auf, die durch Interpolation der aufgezeichneten Messwerte erzeugt wird.

Die generierte Trajektorie wurde erfolgreich an einen Fanuc-Industrieroboter angebunden. Die Reproduktion aufgezeichneter Lackierparameter bei automatisierter Roboterwiedergabe wurde durch reale Sprühversuche validiert und zeigte eine stabile Übertragung der aufgezeichneten Beschichtungsparameter. Die integrierte Schichtdickensimulation aktualisiert das Beschichtungsergebnis nach Bahnänderungen in unter einer Sekunde.

5 Diskussion

Die Sensordatenfusion aus kamerabasiertem Tracking und IMU-Daten erwies sich als essenziell, um Verdeckungssituationen unter realen Bedingungen robust zu beherrschen. Als relevante Randbedingung zeigte sich dabei die Absaugung der Lackierkabine: Eine zu hohe Absaugleistung erzeugt Verwirbelungen, die Sprühnebel zu den Kameras transportieren und die Erkennungsqualität beeinträchtigen. Eine laminare Strömungsführung ist daher eine wichtige konstruktive Anforderung für den Praxiseinsatz. Die Anzahl von acht Kameras stellt einen Kompromiss zwischen Markersichtbarkeit bei Verdeckungen und Systemaufwand dar. Der Kalibrieraufwand bleibt durch den gemeinsamen Kalibrierwürfel begrenzt. Nach dem Verstellen einer Kamera oder einem Wechsel des Targets ist jedoch eine erneute Kalibrierung erforderlich [6]. Die Rechenlast wird durch die Speicherung reiner Markerinformationen statt vollständiger Bilder reduziert. Mit Blick auf die Kosten für kleine und mittlere Unternehmen ist jedoch zu untersuchen, mit wie vielen Kameras die geforderte Genauigkeit noch zuverlässig erreicht wird. Die praxistaugliche Weiterentwicklung des Systems für den direkten Einsatz durch Werker steht im Mittelpunkt künftiger Arbeiten. Die aktuell verwendeten sichtbaren LED-Marker stören den Lackierprozess, weshalb die Ablösung durch Infrarot-LEDs ein konkreter nächster

Schritt ist. Darüber hinaus bleiben die Übertragbarkeit auf weitere Roboterplattformen sowie die Robustheit gegenüber wechselnden Lichtverhältnissen offene Fragen für weiterführende Untersuchungen.

6 Fazit

AuLaTeach zeigt, dass der Wissenstransfer vom erfahrenen Lackierer zum Roboter durch ein markerbasiertes Trackingsystem mit roboterunabhängiger Trajektoriengenerierung technisch realisierbar ist. Der Ansatz adressiert eine bislang wirtschaftlich schwer automatisierbare Fertigungssituation und eröffnet neue Perspektiven für die flexible Beschichtungsautomatisierung bei kleinen Losgrößen. Durch die Trennung von Erfassung und Kinematikanbindung entsteht ein generisch einsetzbares System, das nicht auf eine spezifische Roboterplattform oder ein bestimmtes Beschichtungsverfahren beschränkt ist. Das Prinzip des Vormachens könnte damit als übertragbares Paradigma für weitere handwerklich geprägte Fertigungsprozesse dienen, in denen implizites Erfahrungswissen bislang nur schwer zu automatisieren ist.

Förderhinweis: Dieses Projekt wurde gefördert durch das Bundesministerium für Wirtschaft und Energie (BMWE) aufgrund eines Beschlusses des Deutschen Bundestages (Förderkennzeichen: 16KN117724, Laufzeit: 01.12.2023–30.11.2025). Projektpartner: Gesellschaft zur Förderung angewandter Informatik e.V. und THÜMA Maschinenbau und Service GmbH.

References

[1] P. Merriaux, Y. Dupuis, R. Boutteau, P. Vasseur, and X. Savatier, "A study of Vicon system positioning performance," *Sensors*, vol. 17, no. 7, p. 1591, 2017.

[2] M. Švejda, M. Goubej, A. Jáger, J. Reitinger, and O. Severa, "Affordable motion tracking system for intuitive programming of industrial robots," *Sensors*, vol. 22, no. 13, p. 4962, 2022.

[3] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[4] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnicer, "Generation of fiducial marker dictionaries using mixed integer linear programming," *Pattern Recognition*, vol. 51, pp. 481–491, 2016.

[5] OpenCV contributors, "Charuco detection tutorial." `https://docs.opencv.org/3.4/df/d4a/tutorial_charuco_detection.html`, 2018. Accessed: 2025-01-12.

[6] H. Ben Khedher, "Entwicklung eines Verfahrens zur Posenschätzung mittels Markern und Multikamerasystem," Masterarbeit, Hochschule für Technik und Wirtschaft Berlin, 2025.

[7] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate O(n) solution to the PnP problem," *International Journal of Computer Vision*, vol. 81, no. 2, pp. 155–166, 2009.

[8] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

Entwicklung eines Rückenmodell-Generators zur robusteren Klassifizierung von Skoliose

P. Seelig[1], H. Ben Khedher[2]
[1]Gesellschaft zur Förderung angewandter Informatik e.V., Berlin, seelig@gfai.de
[2]Gesellschaft zur Förderung angewandter Informatik e.V., Berlin, ben_khedher@gfai.de

Abstract: Im Rahmen des Projekts „FlexiBrace – Modellbasierter Entwurf und Fertigung funktionsintegrierter Skoliose-Korsette für das therapiebegleitende Monitoring sowie zur Formanpassung“ wird ein abstrahiertes, parametrisches Rückenmodell entwickelt, das als Grundlage für die modellgestützte Diagnose zur robusten Erkennung und Klassifikation skoliotischer Deformitäten dient. Der vorliegende Beitrag beschränkt sich auf die Modellgenerierung; die Klassifikation selbst ist Gegenstand nachfolgender Arbeiten. Die Generierungsparameter des Modells werden aus nicht-invasiv erfassten Torsodaten der Patienten bzw. Rückenscans abgeleitet, sodass für jeden Patienten ein skaliertes, abstrahiertes Modell eines gesunden Rückens erzeugt werden kann. Das Modell berücksichtigt dabei ausschließlich anatomisch relevante, äußerlich sichtbare beziehungsweise palpierbare Strukturen: das Kreuzbein, die Schulterblätter sowie die Wirbelsäule. Die Längenverhältnisse der einzelnen Wirbelsäulensegmente zur Gesamthöhe der Wirbelsäule wurden auf Basis anatomischer Maßangaben aus der einschlägigen Fachliteratur bestimmt und in das Modell integriert. Das volumetrische Modell wurde anschließend mittels „Bottom-to-Top“ Flexionssimulationen auf grundsätzliche Simulierbarkeit geprüft; die Ergebnisse zeigen, dass eine FEM-Simulation grundsätzlich möglich ist.

Keywords: Parametrisches Rückenmodell, Finite-Elemente-Methode (FEM), Netzgenerierung, 3D-Modellierung, Flexionssimulation

1 Einleitung

Skoliose beschreibt eine dreidimensionale Deformität der Wirbelsäule mit einer dauerhaft bestehenden seitlichen Verkrümmung von mindestens zehn Grad Cobb-Winkel [1, 2]. Die adoleszente idiopathische Skoliose (AIS) ist mit einer Prävalenz von ca. 1–2 % die häufigste Form [1, 2]. Die konservative Behandlung stützt sich gegenwärtig überwiegend auf monolithische Korrekturorthesen (im Folgenden als Korsett bezeichnet), die keine objektive Rückmeldung über Tragzeit und Korrekturwirkung liefern. Zugleich sind etablierte bildgebende Diagnoseverfahren wie Röntgen oder Computertomographie mit ionisierender Strahlung verbunden und daher für eine begleitende Therapieüberwachung nur eingeschränkt geeignet. Diese Limitierungen motivieren im Rahmen des vorwettbewerblichen Forschungsvorhabens FlexiBrace die Entwicklung eines Systems, das ausgehend von strahlungsfreien Rücken- bzw. Torsoscans einen individuellen Korsettvorschlag ableitet und diesen mit form- und sensorintegrierten Komponenten kombiniert, um eine therapiebegleitende Datenerfassung sowie adaptive Formanpassungen zu ermöglichen. Aus den Scandaten sollen dazu der Cobb-Winkel, eine Klassifikation der Deformität nach Rigo und Chêneau und die erforderlichen Korrekturkräfte bestimmt werden. Die Klassifikation ordnet die Krümmungsmuster der Wirbelsäule definierten Typen zu, aus denen sich das passende Korsettdesign ableiten lässt [3]. Voraussetzung hierfür ist ein simulierbares Modell der anatomischen Rückenstruktur, das aus nicht-invasiv ermittelten Parametern generiert werden kann. Ein solches parametrisches Modell ist uns in einer für diese Anwendung geeigneten Form nicht bekannt. Der vorliegende Beitrag fasst die Ergebnisse einer im Rahmen von FlexiBrace durchgeführten Bachelorarbeit zusammen, die diese Lücke durch die Entwicklung eines abstrahierten, parametrischen Rückenmodells für die Finite-Elemente-Methode (kurz FEM) adressiert. Das Modell bildet eine theoretisch gesunde Wirbelsäule ab und dient als Referenzzustand. Es wird patientenindividuell aus den Torsodaten generiert und anschließend soll es an die im Scan erkannten Landmarken deformiert werden. Aus der Abweichung zwischen Referenzmodell und deformiertem Modell sollen der Cobb-Winkel bestimmt und über das inverse Deformationsproblem die der Verkrümmung zugrunde liegenden Kräf-

te approximiert werden. Auf diesen Kräften soll die Auslegung der Korrekturwirkung des Korsetts beruhen, die die Wirbelsäule in Richtung des gesunden Referenzzustands führt. Der Beitrag legt hierfür die Grundlage und konzentriert sich auf die parametrische Geometriegenerierung. Gezeigt wird, dass sich das Modell mittels FEM simulieren lässt. Die Ableitung der Korrekturkräfte sowie die Entwicklung und Validierung eines Klassifikationsverfahrens sind nicht Gegenstand dieser Arbeit, sondern Teil nachfolgender Arbeiten im Projekt. Abschnitt 2 beschreibt die Geometrieerzeugung und die Vernetzung, Abschnitt 3 prüft anhand einer einfachen FEM-Simulation die grundsätzliche Simulierbarkeit, und Abschnitt 4 fasst die Ergebnisse zusammen und gibt einen Ausblick.

2 Abstrahierte anatomische Rückenmodellierung

Die Abstraktion erfolgt gezielt in Abhängigkeit von der jeweiligen Funktion einer Struktur im Modell. Anatomisch charakteristische, äußerlich identifizierbare Landmarken – wie Kreuzbein und Schulterblätter – dienen primär der Referenzierung und Positionierung des Modells relativ zum Torso und werden daher durch einfache geometrische Grundkörper hinreichend genau beschrieben. Die Wirbelsäule selbst wird ohne fortsatzartige Strukturen als Abfolge elliptischer Querschnitte abstrahiert, die zu einem kontinuierlichen Volumenkörper verbunden werden. Für die angestrebte Erkennung und Klassifikation äußerer Deformitätsmuster ist in erster Linie die globale Krümmungsgeometrie der Wirbelsäule ausschlaggebend und nicht die Detailgeometrie einzelner Knochenstrukturen.

Als Eingabedaten dienen strahlungsfrei erfasste Oberflächendaten in Form einer 3D-Punktwolke des Patientenrückens [4] (siehe Abbildung 1). Aus dieser Punktwolke werden die geometrischen Bounding-Box-Parameter (Höhe, Breite, Tiefe) extrahiert. Dazu wird eine Oriented Bounding Box (kurz OBB) angelegt und deren Seitenlängen werden ausgelesen, wie beispielhaft in Abbildung 2 dargestellt. Im vorliegenden Beitrag wird das Modell ausschließlich anhand dieser Maße skaliert; die Anpassung an erkannte Landmarken ist Gegenstand nachfolgender Arbeiten (siehe Abschnitt 4). In den nachfolgenden Abschnitten werden die Generierungsalgorithmen der einzelnen Modellbestandteile beschrieben. Das Modell wird dabei mit den in Abbildung 2 ausgelesenen Dimensionen ($47 \times 36 \times 11$ cm) generiert.



Abbildung 1: Punktwolke eines Patientenrückenscans.

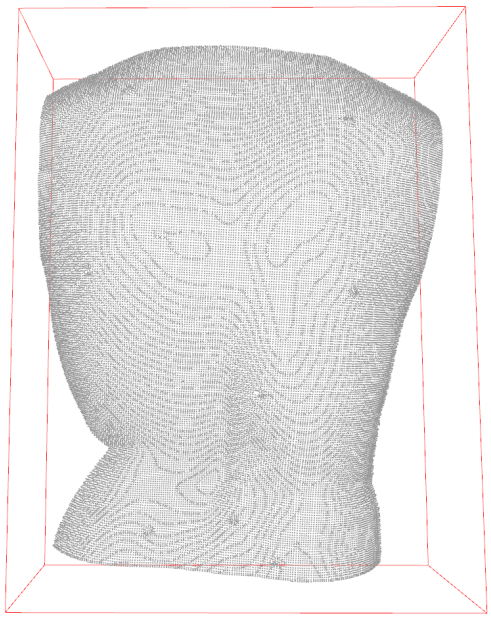


Abbildung 2: Die Oriented Bounding Box um die Punktwolke mit den Dimensionen 47x36x11 cm.

Das Kreuzbein

Das Kreuzbein wird abstrahiert als gerades Prisma mit dreieckiger Grundfläche modelliert, das durch Extrusion entlang der Sagittalachse (y) erzeugt wird. Die Breite der Grundfläche entspricht dem Breitenradius der Ellipse des Wirbels L5, um einen konsistenten Übergang zum letzten Lendenwirbel zu gewährleisten. Die Generierung wird im Algorithmus 1 erläutert.

Eingabe: OBB-Höhe h, L5 w, Tiefe d
Ausgabe: Kreuzbein P

1: $v_r \leftarrow (w,\ 0,\ 0)$
2: $v_l \leftarrow (-w,\ 0,\ 0)$
3: $v_t \leftarrow (0,\ 0,\ -0.15h)$
4: $B \leftarrow \textsc{ErstelleDreieck}(v_r, v_l, v_t)$
5: $P \leftarrow \textsc{Extrudieren}(B, -d)$
6: **Zurückgeben** P

Algorithmus 1: Kreuzbeingenerierung

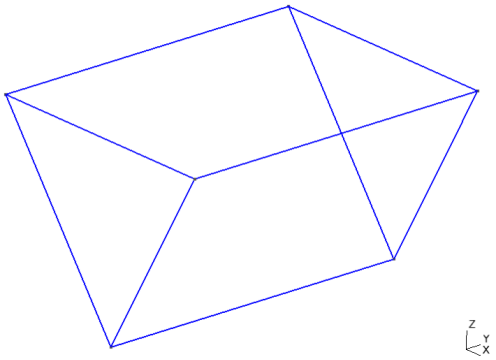


Abbildung 3: Geometrische Darstellung des Kreuzbeines (Tiefe $d = 11\,\text{cm}$).

Da das Kreuzbein nur als geometrischer Grundkörper abstrahiert wird und lediglich der Lagerung des Modells dient, werden die Literaturwerte als Näherung verwendet. Die Höhe des Prismas wird mit $0{,}15\,h$ angesetzt; der Wert ergibt sich aus dem Verhältnis der mittleren Kreuzbeinlänge [5] zur Wirbelsäulenlänge nach [6, Tab. 3]. Geschlechtsspezifische Unterschiede werden dabei nicht berücksichtigt. Die Länge der Extrusion entspricht der Tiefe der Bounding Box und beträgt für das in Abbildung 3 dargestellte Beispielmodell $d = 11\,\text{cm}$.

Die Wirbelsäule

Die Abmessungen der einzelnen Wirbel orientieren sich an den Maßen aus [7, Abb. 4] und [6, Tab. 3]. Daraus werden die Halbachsen jedes Wirbels als Anteil der Wirbelsäulenlänge h abgeleitet, sodass sich die Geometrie konsistent an unterschiedliche Körpergrößen anpasst. Jeder Wirbel wird dabei als Ellipse abstrahiert, deren Halbachsen in Abhängigkeit von der OBB-Höhe h, welche als Wirbelsäulenlänge angenommen wird, bestimmt werden. Die Ellipsen werden jeweils an der Position der Endplatte entlang einer Spline positioniert, welche den Verlauf der Wirbelsäule beschreibt. Da die patientenspezifische Rekonstruktion der Wirbelsäulenkrümmung weiterhin Gegenstand aktueller Forschung ist, wird in der vorliegenden Arbeit eine eigens konstruierte Spline verwendet, die den zu erwartenden Verlauf einer Wirbelsäule approximiert.

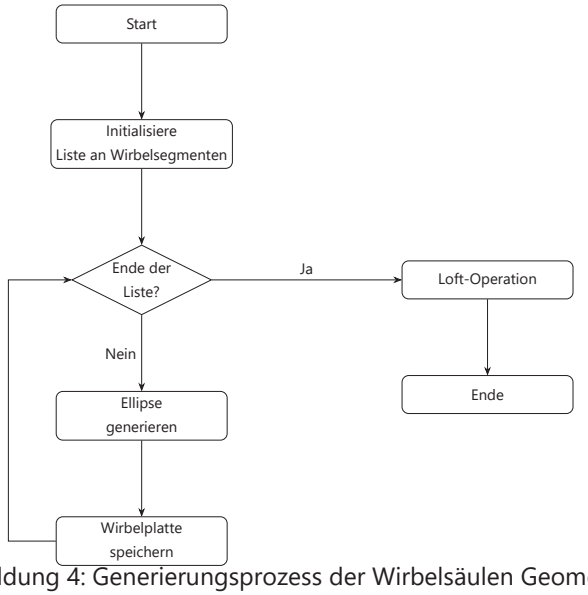


Abbildung 4: Generierungsprozess der Wirbelsäulen Geometrie

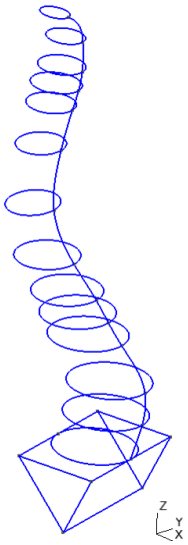


Abbildung 5: Geometrische Darstellung der Wirbelsäule mit Kreuzbeingeometrie (OBB-Höhe $h = 47$ cm).

Abbildung 4 zeigt den grundsätzlichen Ablauf der Erzeugung der Wirbelsäulengeometrie. Zunächst werden die Querschnitte der einzelnen Wirbel entsprechend ihrer Position und ihrer skalierten Abmessungen erzeugt. Anschließend werden diese Querschnitte durch eine Loft-Operation zu einem kontinuierlichen Volumenkörper verbunden (siehe Abbildung 5).

Die Schulterblätter

Die Schulterblätter werden aus einer dreieckigen Grundfläche erzeugt. Zwei Eckpunkte liegen auf der Höhe von dem ersten Brustwirbel und definieren die obere Kante der Grundfläche über die gesamte Schulterblattbreite, welche der Hälfte der OBB-Breite entspricht. Der dritte Eckpunkt wird auf der Höhe des siebten Brustwirbels positioniert und zur Approximation der charakteristischen, nach medial spitz zulaufenden Form des Schulterblatts leicht in x-Richtung versetzt. Dadurch entsteht eine annähernd dreieckige Grundfläche. Zur Erzeugung des Volumens wird die Grundfläche entlang der y-Achse um die Breite des ersten Brustwirbels extrudiert. Das zweite Schulterblatt wird anschließend durch Kopieren der erzeugten Geometrie und eine Rotation um 180° um die z-Achse erzeugt. Dieses Vorgehen gewährleistet eine symmetrische Anordnung beider Schulterblätter bei minimalem Modellierungsaufwand (siehe Abbildung 6).

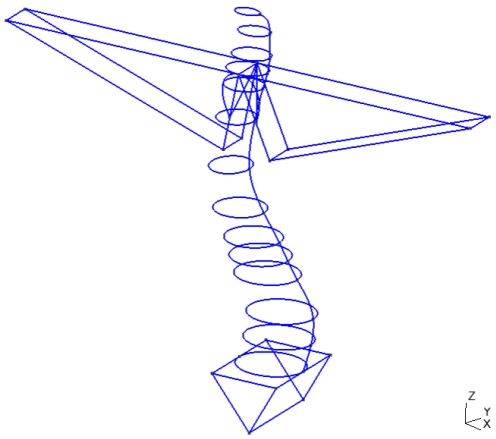


Abbildung 6: Geometrische Darstellung der Schulterblätter (Breite $b = 18$ cm).

Netzgenerierung und Labeling

Die Netzgenerierung erfolgt mit Gmsh, einem Open-Source-FEM-Netzgenerator mit integrierter CAD-Engine, der über GUI, eine eigene Skriptsprache (`*.geo`) oder Programmierschnittstellen für C++, C und Python angesteuert werden kann [8]. Die Modellkomponenten werden zunächst zu einer konsistenten Gesamtgeometrie mit gemeinsamen Grenzflächen verschmolzen (siehe Abbildung 7) und anschließend mit dem Delaunay-Algorithmus (`Mesh.Algorithm3D = 1`) in ein Tetraedernetz mit Elementgrößen von 0,1–0,5 cm überführt. Alle Elemente werden lagebasiert gelabelt, was Kreuzbein, Wirbelsäule und Schulterblätter unterscheidbar macht und die Zuweisung von Materialien, Randbedingungen und Lastflächen ermöglicht. Die Erzeugung des vollständigen Modells inklusive Vernetzung (28 085 Knoten, 158 763 Elemente; Abbildung 7) dauerte auf einem Standard-Desktop-PC (Intel® Core™ i7-8700, 32 GB RAM) rund 11 Sekunden. Da die Vernetzungsparameter die Ergebnisse maßgeblich beeinflussen, sei betont, dass hier nur eine beispielhafte, für die qualitative Plausibilitätsprüfung ausreichende Vernetzung verwendet wurde.

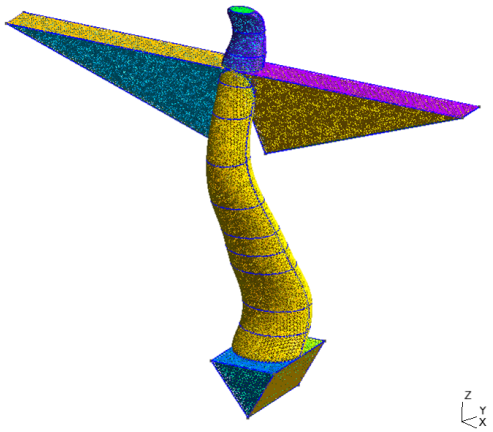


Abbildung 7: Gesamtes Netz des abstrahierten Modells des theoretisch gesunden Rückens.

3 Simulation und Diskussion

Zielsetzung und Methodik

Die folgenden Simulationen sollen zeigen, dass eine FEM-Simulation mit dem in Abschnitt 2 entwickelten Modell grundsätzlich durchführbar ist. Dazu wird die Wirbelsäule segmentweise von unten nach oben (Bottom-to-Top) als einseitig eingespannter Balken untersucht: Das lumbale Segment wird am Kreuzbein eingespannt und an seinem oberen Ende belastet, das thorakale und das zervikale Segment werden analog mit eingespanntem unteren Ende betrachtet. Erwartet wird ein Kragarmverhalten: Innerhalb jedes Segments soll die Verschiebung vom eingespannten zum belasteten Ende hin zunehmen. Ein Vergleich der Verschiebungen zwischen den Segmenten ist nicht Ziel, da die Gesamtkräfte aufgrund der unterschiedlichen Lastflächen voneinander abweichen.

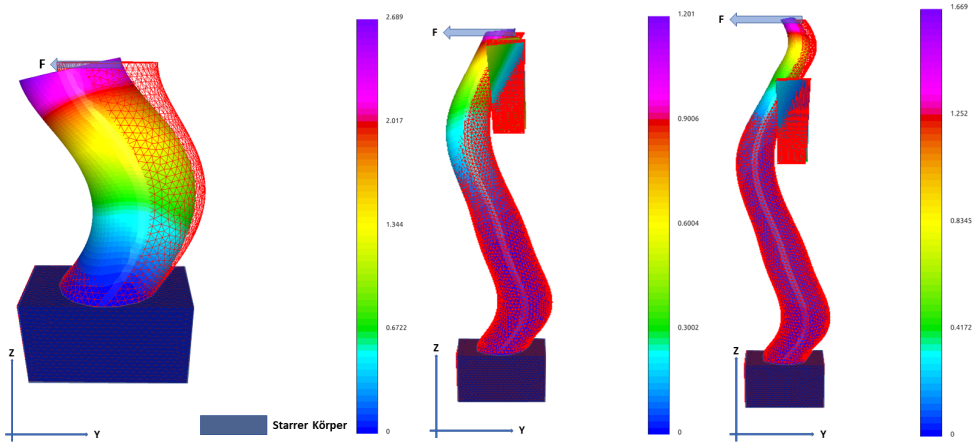
Als Solver dient *MFEM* (*Modular FEM*), eine Open-Source-C++-Bibliothek zur Lösung partieller Differentialgleichungen [9–11]. Das Material wird isotrop und linear-elastisch angenommen; die Lamé-Konstanten berechnen sich aus E-Modul E und Poissonzahl ν:

$$\mu = \frac{E}{2(1+\nu)}, \qquad \lambda = \frac{E\,\nu}{(1+\nu)(1-2\nu)} \tag{1}$$

Für die Wirbelsäule wurden die Werte der Spongiosa verwendet ($E = 100\,\text{MPa}$, $\nu = 0{,}2$, also $\lambda = 27{,}778\,\text{MPa}$ und $\mu = 41{,}667\,\text{MPa}$), da das Wirbelinnere den größten Volumenanteil ausmacht. Sie liegen im literaturüblichen Bereich für humanes Knochengewebe (Spongiosa: einige zehn bis wenige hundert MPa; Kortikalis: 10–20 GPa) [12]. Kreuzbein und Schulterblätter dienen lediglich der Einbettung und Lagerung und wurden um den Faktor 50 versteift, um sie näherungsweise als starr abzubilden. Auf die oberste Deckfläche des jeweiligen Segments wirkt eine konstante Flächenlast von $-0{,}5\,\text{N/mm}^2$ in y-Richtung.

Ergebnisse und Diskussion

Das lumbale Segment verschob sich um 2,7 cm, das thorakale um 1,2 cm und das zervikale um 1,7 cm (Abbildung 8).



(a) Lumbal – Flexion (b) Thorakal – Flexion (c) Zervikal – Flexion

Abbildung 8: Verformung der segmentweisen FEM-Simulation in cm (Rot: belastungsfreier Zustand).

Das erwartete Kragarmverhalten ist in allen drei Simulationen qualitativ erkennbar (Abb. 8). Ein Vergleich zwischen den Segmenten ist nicht möglich, da aufgrund der unterschiedlichen belasteten Flächen unterschiedliche Gesamtkräfte wirken. Die Ergebnisse sind daher ein qualitativer Funktionsnachweis der Simulationspipeline. Zudem enthält das Modell weder Bandscheiben, Rippen, Bänder noch Muskulatur und nutzt für die gesamte Wirbelsäule homogenes Spongiosamaterial, sodass die Verschiebungen nur qualitativ zu interpretieren sind.

4 Fazit und Ausblick

Im Rahmen dieses Beitrags wurde ein parametrisches Rückenmodell entwickelt und in einer segmentweisen Simulation auf grundsätzliche Simulierbarkeit geprüft. Es ist grundsätzlich mittels FEM simulierbar; eine Verifikation und Validierung steht noch aus. Im weiteren Verlauf von FlexiBrace soll das Modell zunächst verifiziert werden, durch einen Vergleich mit der Balkentheorie bei identischer resultierender Kraft und einen systematischen Vergleich von Vernetzungsstrategien, da die Vernetzung die Ergebnisse maßgeblich beeinflusst. Anschließend soll es anhand von Literaturwerten zulässiger Bewegungsumfänge validiert werden. Darauf aufbauend soll das Modell in der Punktwolke des Patientenscans platziert und passend zu den erkannten Landmarken deformiert werden. Aus dem deformierten Modell soll mithilfe von Projektionen der Cobb-Winkel ausgelesen und zur Validierung mit Patientendaten verglichen werden. Ergänzend soll mittels Plane Fitting an die Ellipsen des Modells die axiale Rotation der thorakalen und lumbalen Sektion relativ zum Kreuzbein bestimmt werden. Über das inverse Deformationsproblem sollen die der Deformation zugrunde liegenden Kräfte approximiert werden. Auf Grundlage dieser Kräfte und der Rigo-Klassifikation soll abschließend ein Korsett vorgeschlagen werden.

5 Danksagung

INDUSTRIELLE GEMEINSCHAFTSFORSCHUNG IGF

Die in diesem Beitrag vorgestellten Arbeiten entstanden im Rahmen des Forschungsprojekts „FlexiBrace – Modellbasierter Entwurf und Fertigung funktionsintegrierter Skoliose-Korsette für das therapiebegleitende Monitoring sowie zur Formanpassung (FlexiBrace)“. Das Forschungsvorhaben „FlexiBrace“ der GFal (Gesellschaft zur Förderung angewandter Informatik e.V.), Berlin, wird im Rahmen des Programms zur Förderung der industriellen Gemeinschaftsforschung (IGF) vom Bundesministerium für Wirtschaft und Energie aufgrund eines Beschlusses des Deutschen Bundestages gefördert (Förderkennzeichen 01IF23429N). Die Autoren danken für diese finanzielle Unterstützung.

Literatur

[1] F. Altaf, A. Gibson, Z. Dannawi, and H. Noordeen, "Adolescent idiopathic scoliosis," *BMJ British medical journal*, vol. 346, p. f2508, 2013.

[2] P. Trobisch, O. Suess, and F. Schwab, "Die idiopathische Skoliose," *Dtsch Arztebl Int*, vol. 107, pp. 875–84, 2010.

[3] M. D. Rigo, M. Villagrasa, and D. Gallo, "A specific scoliosis classification correlating with brace treatment: description and reliability," *Scoliosis*, vol. 5, no. 1, p. 1, 2010.

[4] M. Kaiser, "PCdareSoftware," 2024. Accessed: 2026-08-10.

[5] N. Sultana, S. Mannan, and M. Iqbal, "A study of sacral index for identification of sexual dimorphism," *Mymensingh Medical Journal*, vol. 27, no. 4, pp. 710–714, 2018.

[6] I. Busscher, J. J. W. Ploegmakers, G. J. Verkerke, and A. G. Veldhuizen, "Comparative anatomical dimensions of the complete human and porcine spine," *European Spine Journal*, vol. 19, no. 7, pp. 1104–1114, 2010.

[7] C. Vergari, M. Karam, R. Pietton, R. Vialle, I. Ghanem, W. Skalli, and A. Assi, "Spine slenderness and wedging in adolescent idiopathic scoliosis and in asymptomatic population: an observational retrospective study," *European Spine Journal*, vol. 29, 2020.

[8] C. Geuzaine and J.-F. Remacle, "Gmsh: A three-dimensional finite element mesh generator with built-in pre- and post-processing facilities," *International Journal for Numerical Methods in Engineering*, vol. 79, no. 11, pp. 1309–1331, 2009.

[9] "MFEM: Modular finite element methods [Software]." `mfem.org`.

[10] R. Anderson, J. Andrej, A. Barker, J. Bramwell, J.-S. Camier, J. Cerveny, V. Dobrev, Y. Dudouit, A. Fisher, T. Kolev, W. Pazner, M. Stowell, V. Tomov, I. Akkerman, J. Dahm, D. Medina, and S. Zampini, "MFEM: A modular finite element methods library," *Computers & Mathematics with Applications*, vol. 81, pp. 42–74, 2021.

[11] J. Andrej, N. Atallah, J.-P. Bäcker, J.-S. Camier, D. Copeland, V. Dobrev, Y. Dudouit, T. Duswald, B. Keith, D. Kim, T. Kolev, B. Lazarov, K. Mittal, W. Pazner, S. Petrides, S. Shiraiwa, M. Stowell, and V. Tomov, "High-performance finite elements with MFEM," *The International Journal of High Performance Computing Applications*, vol. 38, no. 5, pp. 447–467, 2024.

[12] M. Morimoto, S. Tripathi, M. Kodigudla, E. Motohashi, J. Fujitani, V. K. Goel, and K. Sairyo, "Biomechanical effects of thoracic flexibility and stiffness on lumbar spine loading: A finite element analysis study," *World Neurosurgery*, vol. 184, pp. e282–e290, 2024.

Photogrammetric 3D Model-Based Animal Behavior Classification

Hellwich, Olaf[1,5], Heine, Moritz[1], Reiske, Patrik[1,5],Traverso, Sole[3,1,5], Hohlbaum, Katharina[2,5], Sprekeler, Henning[3,5], Thöne-Reineke, Christa[4,5], Lewejohann, Lars[2,4,5], Bhuiyan, Roman[1]
[1]TU Berlin, Computer Vision & Remote Sensing
[2]German Centre for the Protection of Laboratory Animals (Bf3R), German Federal Institute for Risk Assessment (BfR)
[3]TU Berlin, Modeling of Cognitive Processes
[4]FU Berlin, Institute of Animal Welfare, Animal Behavior and Laboratory Animal Science
[5]TU Berlin, Science of Intelligence Excellence Cluster

Abstract: Video-based behavior estimation is often implemented end-to-end from image sequences. We advocate a two-stage alternative: reconstruct 3D body and object geometry before classifying behavior. This makes ambiguities of the ill-posed 2D-to-3D inverse problem explicit and potentially correctable. We demonstrate the strategy with simulated interacting mice and extend it to real video.

Keywords: Video-based animal observation; 3D motion reconstruction; Behavior classification; Behavior simulation.

1 Introduction

Behavior is a higher-level construct inferred from measurable movement patterns. In a semantic network, video features may connect directly to movement, whereas behavior is inferred indirectly through movement [1]. Because video provides only 2D projections and cannot capture all causal subtleties of living behavior, behavior is not directly observable; inferring it from video is therefore an ill-posed inverse problem. We therefore propose a fully three-dimensional (3D) simulation and reconstruction framework: 3D body kinematics are reconstructed from video first, and behavioral states are inferred afterward rather than directly from raw frames.

Reconstruction also uses prior 3D models: deformable animal bodies, static cage geometry, and kinematic enrichment devices such as lock-boxes. These provide structural and motion constraints. We illustrate the concept with simulated interactions between two mice and real multi-view video of a mouse solving a lock-box task.

2 State-of-the-Art of Animal Behavior Classification

Modern video-based animal behavior pipelines divide activity into discrete, functionally meaningful classes. They typically track body parts with pose estimators such as DeepLabCut [2], then model behavior from the resulting trajectories using supervised or unsupervised temporal analysis. Traditional systems mainly use monocular 2D keypoints, while newer methods increasingly accept reconstructed 3D trajectories.

MARS recognized mouse social behaviors from video [3]; B-SOiD discovered behavioral motifs unsupervised from pose trajectories [4]; and A-SOiD added active learning to reduce annotation while improving classification [5]. Although commonly demonstrated on 2D data, B-SOiD can in principle incorporate 3D trajectories.

VAME learns latent motion representations with variational sequence models and supports 2D or reconstructed 3D trajectories [6]. SimBA provides supervised, feature-based behavior classification from pose outputs and can likewise accommodate 3D keypoints [7].

Keypoint-MoSeq extends MoSeq to pose-estimation keypoints and naturally supports 2D and 3D skeletons [8]. DLC2Action performs temporal action recognition from DeepLabCut-style trajectories and is similarly compatible with both 2D and 3D settings [9].

3 Example A: 3D Reconstruction and Behavior Classification of Two Interacting Simulated Mice

Two simulated mice exhibit running, rearing, and intermittent social interaction in an enclosed environment. [1]

3.1 Mechanistic Simulation of Two Interacting Mice

Each simulated mouse is an autonomous agent with a planar trajectory, discrete behavior (running, rearing, or social interaction), a keypoint skeleton, and non-overlap collision handling. Its 3D keypoints are projected into virtual camera images. Control is hierarchical: hard physical and geometric constraints take precedence, followed by social objectives and then individual running or rearing. Social interaction therefore overrides individual behavior when necessary.

The body is a 3D keypoint graph whose nodes are anatomical landmarks and whose edges encode skeletal connectivity. Each trajectory stores planar position, heading, speed, and yaw rate; speed and yaw are damped and noise-perturbed, then adjusted for wall repulsion, social attraction, and inter-mouse avoidance. The mouse then advances along its heading, producing smooth stochastic motion with bounded speed and turning.

Individual states are running or rearing, with timers, gait phase, and rearing parameters. Running lasts 2–6 s and rearing 0.6–1.6 s; after running, rearing begins with probability 0.25, while rearing always returns to running. During social episodes, these timers are frozen. Running adds sinusoidal head jitter and a trot-like triangular paw cycle with diagonal pairs in phase; gait phase scales with speed and stride length. Rearing rotates the skeleton around the tail root, fixes the hind paws, and repositions the forepaws. This procedural pose ignores dynamics such as balance, torques, and joint limits.

Collision handling approximates each body by front and rear 2D spheres derived from the keypoint graph. After movement, projection keeps spheres inside the cage and separates overlapping mice, with dominance weights controlling how much each mouse is displaced.

Social interaction is a pair-level finite-state machine with “none”, “approach”, “hold”, and “separate” phases. Its state records the phase, initiator, meeting point, touch time, and minimum nose distance. Episodes are scheduled stochastically with a cool-down; the initiator animal approaches more directly than the responder, close noses trigger holding, and a fixed hold is followed by separation until sufficient distance is reached.

Per frame, arbitration updates the social state, individual timers when non-social, steering, trajectories, hard constraints, local and world skeleton poses, and finally 3D keypoints and ethogram labels. The skeleton is projected into virtual cameras to produce the synthetic video (Fig. 1, top left).

3.2 Video Rendering

Cage surfaces are rendered by homographically transforming textured images. Mouse bodies use a textured mesh deformed by a 3D rubber-sheet transformation from corresponding skeletal and mesh landmarks. Videos are rendered from three orthogonal views (Fig. 1, top right, and bottom row).

3.3 Extraction of 2D Body-Part Keypoints and 2D Body Segments

DeepLabCut (multi-animal maDLC, version 2.3.11, [10]) was trained on arbitrarily generated simulation images from top, front, and side views. This synthetic setting permits essentially unlimited labeled training data. Under mutual occlusion (Fig. 2), extracted skeletons can connect parts from different mice (Fig. 3), while identities may flip over time and disagree across views.

Mouse body segmentation was also learned from rendered masks for background, mouse 0, and mouse 1. A multi-class 2D encoder-decoder U-Net with GroupNorm and previous-mask conditioning was trained [11]. Example predictions for the image in Fig. 5 are shown in Fig. 6.

[1] The code for this example was created from early February 2026 until mid-April 2026 using ChatGPT Plus 5.2 to 5.5. It is made available at git@github.com:hellwich/mouseModelPaper.git including the LLM chat history.

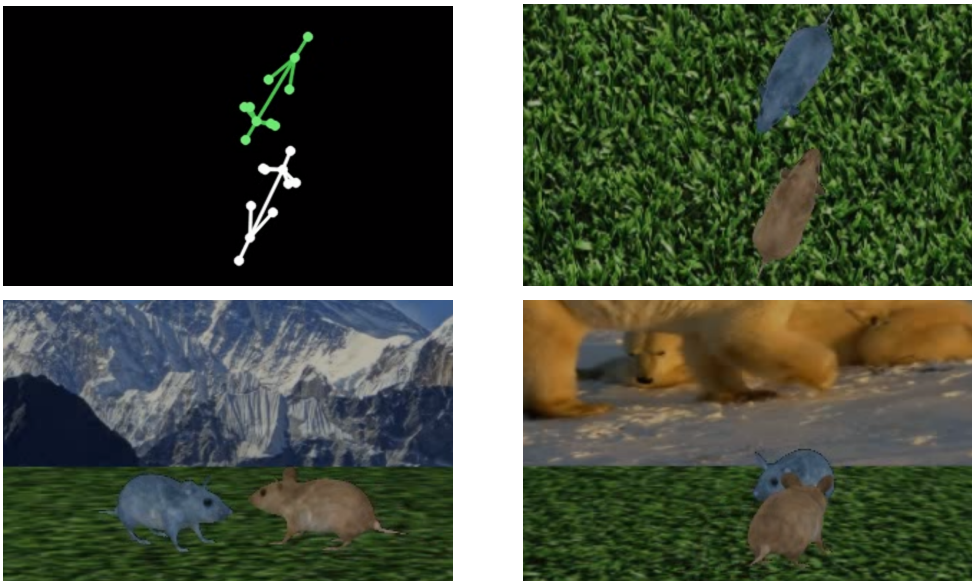


Fig. 1: Simulated skeleton and rendered video. Top, front and side view

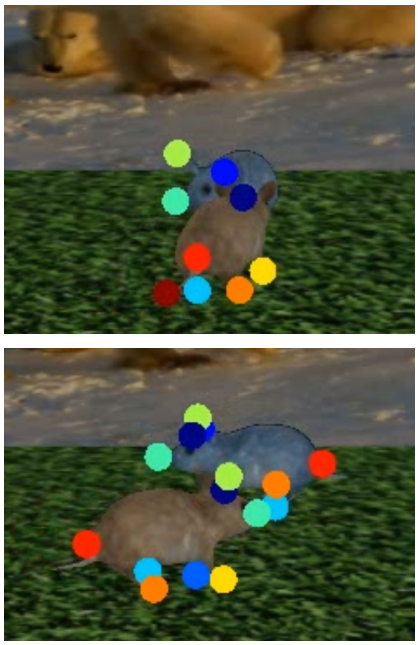


Fig. 2: DLC keypoints

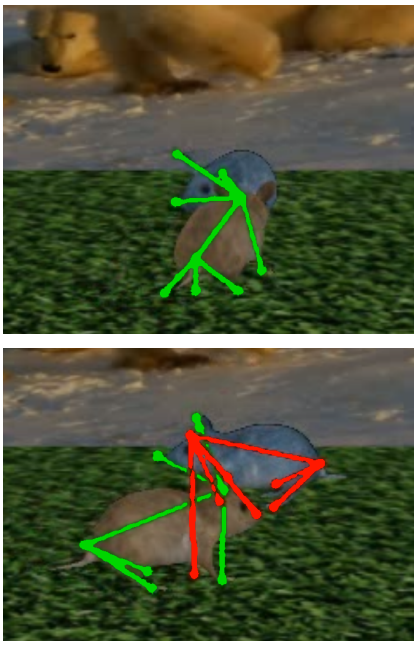


Fig. 3: The same DLC keypoints with skeleton graphs

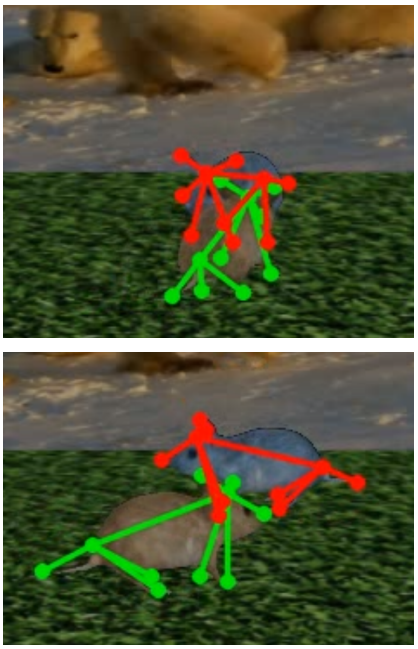


Fig. 4: 3D skeletons after keypoint cleaning



Fig. 5: Overlapping 2D body segments



Fig. 6: U-net semantic segmentation of overlapping bodies

3.4 3D Reconstruction From Multi-View 2D Observations, 3D-Model and Temporal Constraints

Multi-stage reconstruction combines calibrated cameras, per-frame 2D keypoints, and segmentation masks to produce cleaned 3D trajectories and corrected identities for both mice.

For every body part and frame, all plausible two- and three-camera detection combinations are tested, regardless of their original mouse labels. Identity inconsistencies across views are therefore resolved by selecting combinations that are geometrically plausible in 3D.

Stage A validates 3D hypotheses from calibrated viewing rays. Candidate pairs or triplets must have sufficiently non-degenerate intersection angles, pass near one another in 3D, and reproject accurately into the source images. Accepted points are refined by Ceres single-point bundle adjustment [12], with plausible three-view triplets preferred over pairs. A linearized least-squares extension can additionally impose a deformable-mouse motion-track constraint [13]. Per body part, up to two disjoint 3D hypotheses are then selected, so no 2D detection is allowed to contribute to both mice. Untriangulated detections remain as single-view observations for later assignment.

Stage B assigns accepted 3D points and remaining single-view detections to mouse 0 or mouse 1 using multiple identity cues. First, semantic segmentation provides local identity votes for each 2D keypoint. Presence, rather than pixel area size, determines support; so a small visible ear can identify the foreground mouse even when a larger occluded body lies behind it. Votes are accumulated across contributing views, and exclusive support yields an identity cue. Second, temporal continuity repairs frame-to-frame identity flips.

Third, assigned head and tail-root anchors define a centroid and body axis for each mouse. Other 3D parts are assigned by a cost combining segmentation exclusivity, temporal continuity, distance to the centroid, and distance to the body axis, allowing geometric evidence to override the original DLC identity.

Fourth, the prior 3D skeleton checks template-relative edge lengths and alternative head/tail-root pairings to prevent chimeric "half-stitch" bodies combining parts of different mice. Violating skeleton leaf parts may be freed, reassigned, or dropped, while rigid template fitting provides a fallback plausibility test when enough parts are available. Finally, remaining single-view detections are assigned.

This cleaning yields 3D skeleton tracks comparable to the simulated ground truth, although still incomplete and imperfect. Fig. 4 shows results for the preceding DLC examples.

3.5 Behavior Classification from 3D Body-Part Tracks

Both 3D skeleton tracks are converted to normalized, pair-centered temporal features used by sequence models to predict interaction and per-animal rearing, producing a framewise ethogram.

Coordinates are expressed in a local pair-centered co-ordinate system defined by the tail roots, with x directed from mouse 0 to mouse 1 and z vertical, reducing sensitivity to global position and rotation. Velocities and nose-to-nose and tail-root distances are added.

The classification net uses three binary heads: interaction, rearing of mouse 0, and rearing of mouse 1. LSTM, xLSTM, and Transformer models operate on sliding temporal windows rather than isolated frames. Each uses a sequence backbone followed by the same three binary behavior heads.

Table 1 reports F1 scores in a three-minute test sequence. The classification of reconstructed 3D tracks closely matches the classification of simulated ground-truth tracks, suggesting that the reconstruction pipeline preserves much of the information needed to distinguish evaluated behaviors despite ambiguous and incomplete 2D observations.

An end-to-end Transformer trained on perfect simulated multiview 2D projections, using pair-centered coordinates, performs substantially worse when evaluated on DLC keypoints from the same test sequence. Poor rearing performance also persists when this model receives perfect simulated test projections, indicating that DLC detection and identity errors alone do not explain the gap. These results support the effectiveness of the explicit 3D-reconstruction-based pipeline in the investigated setting but do not establish that 3D reconstruction is necessary or that a better-designed end-to-end model could not achieve comparable performance.

Model	Input source	Interaction F1	Rear0 F1	Rear1 F1
LSTM	Simulated 3D tracks	0.6276	0.8369	0.8316
	Reconstructed 3D tracks	0.6147	0.8116	0.8066
xLSTM	Simulated 3D tracks	0.7146	0.7500	**0.8866**
	Reconstructed 3D tracks	0.6702	0.6776	0.8293
Transformer	Simulated 3D tracks	**0.7271**	**0.8611**	0.8743
	Reconstructed 3D tracks	**0.7240**	**0.8286**	**0.8432**
	End-to-end multiview DLC 2D keypoints	0.4071	0.0960	0.1351

Table 1: F1 scores for behavior classification on the same three-minute simulated test sequence. The 3D classifiers were trained on simulated 3D body-part tracks and evaluated on simulated or cleaned, completed 3D-reconstructed tracks. The direct Transformer was trained on perfect multiview 2D projections of the same behavior-training trajectory, using the same temporal training/validation split, and evaluated on DLC 2D detections. Its inputs use pair-centered image coordinates, without segmentation, explicit 3D reconstruction, or additional identity correction. Rearing scores are evaluated on ground-truth non-interaction frames. Bold values denote the highest score for each behavior within each of the two 3D input-source groups.

4 Example B: Empirical Observation of a Live Mouse Performing a Lock-Box Problem-Solving Task

The second example is not yet fully evaluated with the two-stage pipeline; it instead illustrates weaknesses of video behavior classification without intermediate 3D reconstruction.

The lock-box task requires rotating a lever, removing a splint, moving a ball, and shifting a lid to reach an oat flake reward. The data set of [14] provides DeepLabCut-extracted body and lock-box keypoints [2]; despite methodological differences from Example A, similar cleaning is required (Fig. 7).

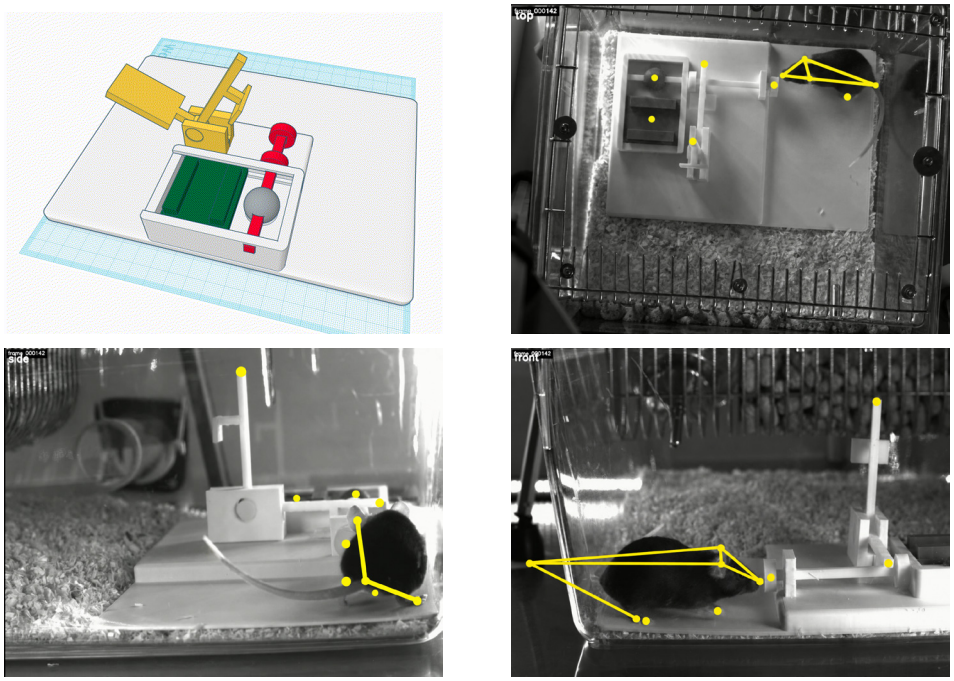


Fig. 7: Mouse lock box experiment. The tail root in the front view is wrongly extracted.

4.1 Calibrating the Cameras With the Lock-Box

Combining the three views requires the position, orientation, and focal length of each camera. These are usually determined from images of a calibration pattern. Here, the lock-box takes this role, since its exact 3D shape is known from its CAD model. For each camera, the parameters are chosen such that the projected CAD model covers the lock-box in the image (Fig. 8).

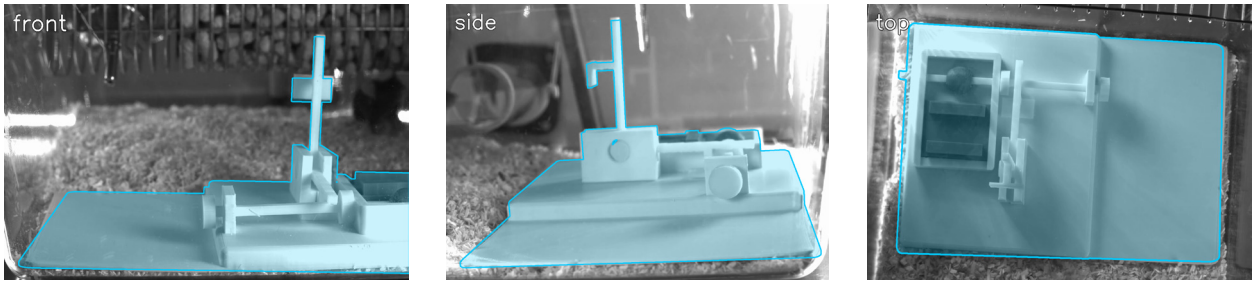


Fig. 8: Camera calibration with the lock-box. The lock-box CAD model (blue), projected with the fitted camera parameters, covers the lock-box in the front, side, and top views.

Fitting each image separately is not sufficient: a single view leaves trade-offs, e.g. between focal length and camera distance, and the three separately fitted cameras disagree on where lock-box landmarks such as the lever tip lie in 3D. The cameras are therefore refined jointly, additionally requiring that the lock-box keypoints detected by DeepLabCut be consistent across views. The measured width of the lock-box base (199 mm) sets the metric scale. The lock-box also defines the coordinate frame, so mouse positions are expressed directly relative to the puzzle.

4.2 From 2D Keypoints to 3D Keypoints

A keypoint detected in one image defines a viewing ray in space; the rays of two cameras intersect at the keypoint's 3D position (triangulation). With three cameras, any two views suffice to compute a 3D point, and the third view provides an independent check: the 3D point is projected into it and compared with the detection there. A point is accepted if all three views agree. If only two views provide reliable detections, the point is accepted but marked as unchecked. If the third view contradicts the other two, the point is rejected, as are points implying implausible bone lengths or jumps between frames. Rejected points are not replaced by guesses; only gaps of up to five frames (0.17 s) are interpolated and marked as such.

Fig. 9(a)–(c) shows 3D keypoints reconstructed in this way and projected back into the three images: in every view they lie on the corresponding body parts.

4.3 Keypoint Errors Revealed by the Third View

Across the recording, roughly a third of all mouse keypoint detections are inconsistent with the other views, and many of these are confident detector errors. In about 40 % of the frames in which DeepLabCut reports the tail root with a likelihood above 0.9 in all three views, the three detections cannot belong to a single 3D point. Typically, the side and top views agree while the front-view detection lies far from the mouse (Fig. 7 and Fig. 9(d)). Within the front view alone, nothing flags such an error, as the detector reports high confidence. A classifier operating directly on 2D keypoints therefore receives these errors as valid input, whereas the 3D reconstruction detects and removes them through the geometric constraint between views.

4.4 Accuracy of the Reconstruction

The real recording has no 3D ground truth, so accuracy cannot be measured directly. We therefore estimated it in a simulation built from the real data: the reconstructed mouse motion is treated as known ground truth, projected into the three calibrated cameras, perturbed with the detection noise and missing detections observed in the real recording, and reconstructed again with the same pipeline. The median 3D error is about 2 mm for head and tail root and about 4 mm for the paws. If the cameras are additionally misaligned by 2°, the errors rise to about 14 mm, so the reconstruction is most sensitive to camera calibration. These results hold for one recording and the assumed noise model.

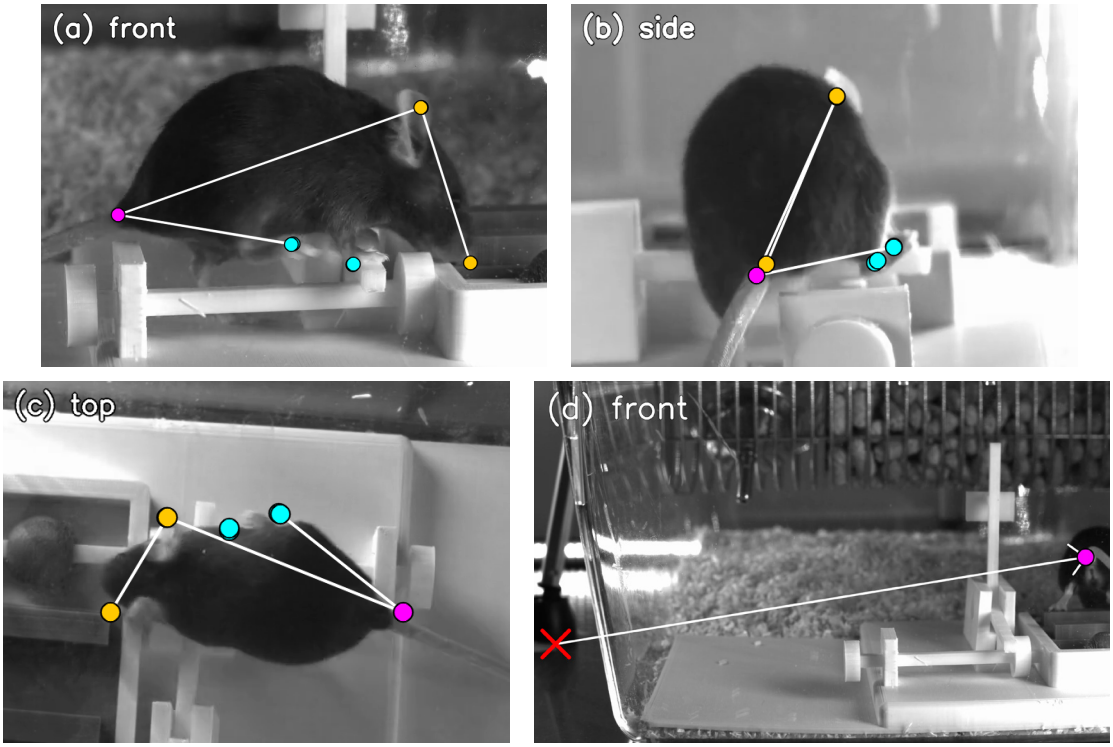


Fig. 9: 3D reconstruction of the mouse. (a)–(c) Keypoints reconstructed in 3D from all three views and projected back into each image: head (nose, ears) ●, paws ●, tail root ●; white lines connect neighboring keypoints of the skeleton. (d) A keypoint error revealed by geometry: DeepLabCut places the tail root at the red cross (likelihood 0.95) at the left image border. Side and top views agree on a tail-root position whose projection into the front view (●) lies on the mouse, about 1,800 pixels away.

5 Conclusion

The examples support three conclusions:

- Preferably, behavior should not be directly estimated from video observation, but first 3D reconstruction should be conducted, and then behavior should be inferred from 3D estimates. This does not exclude replacing explicit 3D reconstruction with implicit neural representation.
- Behavior analysis of real animals profits from parallel full simulation of the animals' behavior, video acquisition, 3D reconstruction, and behavior estimation.
- Incorporating static and dynamic 3D model-based constraints, along with motion constraints, into the traditional image observation constraints enhances 3D reconstruction from video data.

Behavior classification approaches using a single video camera to observe the animals can be integrated into our concept by "depth from single image" networks and geometric 3D models providing constraints for 3D reconstruction.

Following our strategy does not exclude later switching to end-to-end deep-learning strategies that incorporate e.g. implicit 3D representations.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2002/1 "Science of Intelligence" – project number 390523135, and through the Research Training Group GRK 3051-„3RTG".

References

[1] O. Hellwich and C. Wiedemann, "Object extraction from high-resolution multisensor image data," in *Third International Conference Fusion of Earth Data, Sophia Antipolis*, vol. 115, 2000.

[2] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning," *Nature neuroscience*, vol. 21, no. 9, pp. 1281–1289, 2018.

[3] C. Segalin, J. Williams, T. Karigo, M. Hui, M. Zelikowsky, J. J. Sun, P. Perona, and D. J. Anderson, "The mouse action recognition system (MARS) software pipeline for automated analysis of social behaviors in mice," *eLife*, vol. 10, p. e63720, 2021.

[4] A. I. Hsu and E. A. Yttri, "B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors," *Nature Communications*, vol. 12, no. 5188, 2021.

[5] A. I. Hsu and E. A. Yttri, "A-SOiD: active learning for expert-guided discovery of behavior," *Nature Methods*, vol. 21, pp. 597–605, 2024.

[6] K. Luxem, F. Fuhrmann, J. Kürsch, S. Remy, P. Bauer, D. Schubert, and A. A. Kühn, "Identifying behavioral structure from deep variational embeddings of animal motion," *Communications Biology*, vol. 5, no. 600, 2022.

[7] S. R. Nilsson, N. L. Goodwin, J. J. Choong, S. Hwang, H. R. Wright, Z. T. Norville, P. Serreau, S. Peron, and S. A. Golden, "Simple behavioral analysis (SimBA) – an open source toolkit for computer classification of complex social behaviors in experimental animals," *Nature Neuroscience*, vol. 27, pp. 1003–1015, 2024.

[8] C. Weinreb, C. Sosa-Plata, A. Bendesky, and S. R. Datta, "Keypoint-MoSeq: parsing behavior by linking point tracking to pose dynamics," *Nature Methods*, vol. 21, pp. 1146–1156, 2024.

[9] E. Kozlova, A. Bonnetto, and A. Matthis, "DLC2Action: Deep learning for animal action recognition from pose trajectories," *bioRxiv*, 2025. Preprint.

[10] J. Lauer, M. Zhou, S. Ye, W. Menegas, S. Schneider, T. Nath, M. M. Rahman, V. Di Santo, D. Soberanes, G. Feng, *et al.*, "Multi-animal pose estimation, identification and tracking with DeepLabCut," *Nature Methods*, vol. 19, no. 4, pp. 496–504, 2022.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, Springer, 2015.

[12] S. Agarwal, K. Mierle, and T. C. S. Team, "Ceres Solver," 10 2023.

[13] O. Hellwich, N. Andresen, K. Hohlbaum, M. N. Boon, M. Kwiatkowski, S. Matern, P. Reiske, H. Sprekeler, C. Thöne-Reineke, L. Lewejohann, H. Ghani Zada, M. Brück, and S. Traverso, "Tracking mouse from incomplete body-part observations and deep-learned deformable-mouse model motion-track constraint for behavior analysis," *Festschrift für Prof. Dr.-Ing. Helmut Mayer zum 60. Geburtstag*, pp. 45–52, 2024.

[14] M. N. Boon, N. Andresen, S. Traverso, S. Meier, F. Schuessler, O. Hellwich, L. Lewejohann, C. Thöne-Reineke, H. Sprekeler, and K. Hohlbaum, "Mechanical problem solving in mice," *bioRxiv*, pp. 2024–07, 2024.

Automated robust 3D defect reconstruction for additive repair processes

Robert Joost[1], Lukas Engeländer[1,2], Pascal Lünnemann[1]
[1]Fraunhofer Institute for Production Systems and Design Technology IPK, Pascalstraße 8-9, 10587 Berlin, Germany
[2]TU Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany

October 6, 2026

Abstract

This paper proposes an automated process to reconstruct arbitrary 3D defect geometries and their corresponding difference volumes from 3D scan data, enabling adaptive repair processes in Maintenance, Repair and Overhaul (MRO). Currently, creating these defect geometries is a tedious manual task due to highly individual and arbitrary manifestations of defects, and industry-ready automated solutions are lacking. Our approach assumes that an as-designed CAD model is available, allowing not only the reconstruction of defective surfaces but the complete defective volume. The pipeline consists of pre-processing, registration, segmentation, and NURBS approximation. A key element is the robust freeform surface approximation: the defective surface patch is transformed into 2D using the Isomap algorithm, regularized with the Laplace algorithm, and mapped back to 3D to generate a start surface for iterative NURBS fitting. This fitted surface then trims the as-designed CAD model to yield the difference volume. The process is demonstrated with a practical example of a gear cutting tool with a broken tooth, showing its versatility and robustness. Future work will focus on reducing computational effort and integrating the method to handle large industrial parts.

1 Introduction

In a variety of industrial applications, a parametrized representation of the as-is geometry of objects or components is required. In Maintenance, Repair and Overhaul (MRO) activities, adaptive repair processes require precise geometry models to apply additive or subtractive CAM path planning. A typical application of such MRO activities is the repair of high-value components, often made of scarce alloys [1] or produced in relatively small quantities – such as turbine blades [2, 3] or large cutting tools [4]. For such components, the repair process is usually expensive and tedious. However, material scarcity and high initial manufacturing costs make it economically attractive to apply repair instead of new production. In the MRO context, we consider geometry reconstruction of macroscopic, volumetric defects, such as broken areas, spallation, dents, or deformations. To obtain such models, usually the method of 3D model reconstruction (or Reverse Engineering) is applied. The as-is state is captured, typically by 3D scanning. The creation of defect geometries from scan data that serve for repair path planning is often a tedious, manual task. This is caused by highly individual manifestations of defects in practice – shape and location are arbitrary. An automated repair process must be adaptive to this variability. Therefore, a robust and versatile automated geometry reconstruction is crucial to achieve a high degree of automation – at the same time, industry-ready solutions are lacking.

This paper proposes an automated process to reconstruct arbitrary 3D defect geometries, ready to be integrated into adaptive repair processes. We assume that an as-designed CAD model is at hand, enabling the reconstruction of the complete defective volume ("difference volume") to be repaired. The research is embedded in a project that incorporates further parts of the adaptive repair process not described in this work: an integrated inspection system with additional 2D defect detection, subsequent CAM path planning and additive repair using wire-based Directed Energy Deposition (DED). The theoretical background is provided in Section 2. As explained in Section 3, the process consists of typical Reverse Engineering steps: Pre-processing, registration with as-designed CAD, correspondence detection, segmentation, NURBS approximation, and modeling. A special focus is set on the NURBS approximation. Section 4 presents the validation with a gear cutting tool that has a broken tooth. Section 5 concludes that the process is a versatile and robust approach to create such difference volumes. However, it relies on having a precise as-designed CAD model and requires accurate registration of CAD and scan data. Future work is identified with respect to reduction of computational efforts and integration into adaptive repair processes to handle large parts.

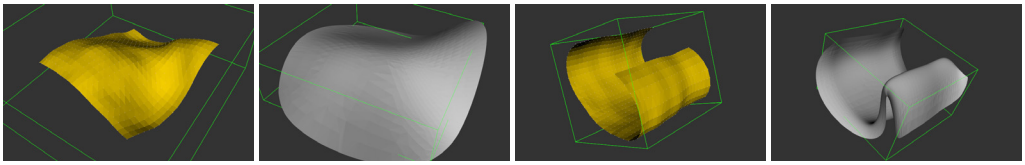
2 Theoretical background

2.1 Adaptive repair processes

MRO processes usually consist of inspection, disposition, execution and verification [5]. The manifestation of occurring defects in terms of shape, location, size etc. can be considered arbitrary. In conventional repair and overhaul processes, this typically requires manual intervention for all steps, since tasks such as defect assessment, measurement, CAM planning, and machine control become highly individual [6]. Adaptive repair processes are primarily aimed at reducing the amount of manual intervention by making use of automated inspection [1]: robotic path planning derived from the captured geometry, camera- and scanner-based defect detection and quantification instead of visual inspection [7], rule-based disposition, consistent digital models and reference systems [8, 9], and integration of inspection directly into the repair facility to allow iterative execution. This work contributes the geometry reconstruction required for the first two aspects.

2.2 NURBS approximation

Since we consider macroscopic defects and their digitization, the method of point cloud approximation by Non-Uniform Rational B-Spline Surfaces (NURBS) [10] is considered. The goal is to approximate a set of points (or additionally, a triangulated point cloud, e.g. a discrete mesh) in $\mathbb{R}^3$ using B-Splines. The standard procedure is to find a parametrization of the underlying points. Zhu et al. [11] group parametrization methods for unorganized point clouds into base-surface projections, meshless parametrization, spherical mappings, mesh-adapted and learning-based schemes. Base-surface methods project the points onto a simple auxiliary surface and inherit its parametrization; they are robust against non-uniform sampling but lose injectivity for markedly non-planar data, so that a subsequent spline fit folds or self-intersects. As an example, Figure 1 shows the application of the *FittingSurface* algorithm provided in the *ON_nurbs* submodule of the *Point Cloud Library* (PCL version 1.15.1 [12]), with a successful approximation of a "planar" region (a,b) and a failed process due to folding (c,d). Meshless parametrization needs no connectivity but relies on a global sparse system, neighbourhood size and boundary detection, whereas spherical, mesh-adapted and learning-based schemes require closed genus-zero shapes, reliable connectivity or training data. As our defect patches are open and strongly curved, it is precisely the projection-based initialization that fails, motivating the embedding-based approach of Section 3.



(a) (b) (c) (d)

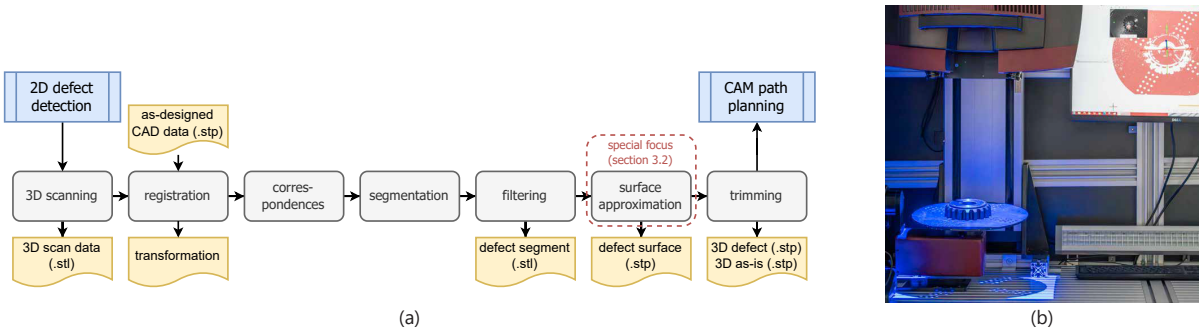
Fig. 1: Approximations with PCA-based embedding (a,b) successful example (c,d) unsuccessful example

3 Concept and Development

3.1 Overall process

The scoped process follows a typical Reverse Engineering procedure and aims at reconstructing the difference volume of a damaged part (Figure 2a). Preceding 2D defect detection and subsequent CAM path planning and repair are not described here. Figure 2b shows the hardware setup, comprising a structured-light *ATOS Triple Scan* scanner and a rotating and tilting table. The processing software is run on a separate PC, linked via REST API. The process starts with 3D scanning, yielding the 3D scan data. The alignment of as-designed and as-is is done via sampling and a multi-level registration approach using Fast Point Feature Histograms (FPFH)-matching and multi-level Iterative Closest Point (ICP). Computing correspondences yields candidates for defect surfaces, which are further segmented and filtered, to finally obtain the actual defect segments. The surface approximation described in Section 3.2 yields

the parametrized defect surface as STEP geometry (.stp), which is then used to trim the as-designed CAD. This results in the parametrized 3D as-is geometry and its counterpart, the difference volume.



(a) (b)

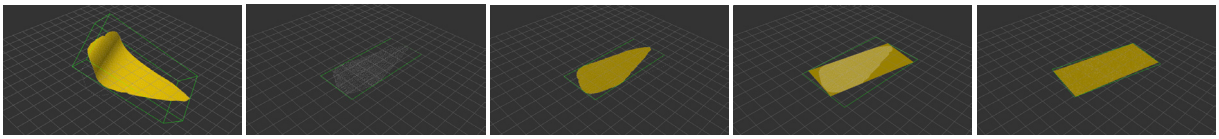
Fig. 2: (a) Overall difference reconstruction process (b) hardware setup

3.2 NURBS Approximation

The proposed approximation follows the two-stage paradigm common to most point cloud parametrization schemes: first embed the data in a planar parameter domain, then fit a spline surface over that domain. It differs from existing work in three respects. First, the parameter domain is not obtained by projecting the points onto an auxiliary surface, but by a geodesic-distance-approximating embedding (Isomap [13]), which remains injective for the strongly curved, non-projectable patches typical of macroscopic defects. Second, the embedding is used not only as the parameter domain but also to construct the initial control point grid, so that the fit starts from an already curved, data-adjacent surface instead of a PCA plane. Third, the parameter values of the input points are known by construction and are passed to the fitting routine directly, which removes the fragile automatic foot-point estimation that fails on curved initial surfaces. The individual steps are described in the following.

To obtain a suitable control point grid in 3D that serves as a start surface for the subsequent approximation, we create a 2D embedding of the point cloud using the Isomap algorithm [13] (scikit-learn implementation [14]). The embedding is then fitted into a rectangle: we compute its smallest enclosing rectangle and snap all boundary vertices onto the rectangle border. Boundaries can either be reconstructed generically from point cloud data, or, if available, by using mesh boundaries. Afterwards the interior vertices are relaxed with the Laplacian mesh smoothing filter provided by VTK [15]. The result is a flattened point cloud (in our case together with a mesh) that covers a rectangular region in the plane and can therefore be used as a parameter domain. Figure 3 illustrates these steps from mesh segment to rectangular parameter region.

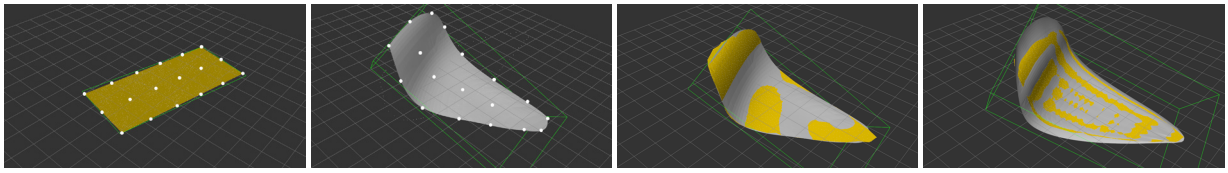
The embedding step gives us a direct correspondence between the embedded mesh and the input mesh: every vertex of the embedded mesh in the parameter region is linked to exactly one vertex of the input mesh in 3D. Because the Point Cloud Library stores meshes as a list of vertices plus connectivity, this correspondence is only known for the vertices, not for arbitrary points of the parameter region. What we need, however, is a mapping that works for any point of the parameter region, so that we can sample it on a regular grid. We therefore extend the known vertex correspondence to the whole parameter region. The straightforward way is to locate the embedded triangle that contains the query point and to interpolate between the 3D positions of its three vertices using barycentric coordinates. This reproduces the surface exactly, but PCL offers no accelerated data structure for point-in-triangle queries. For performance reasons we instead pick the embedded triangle whose centroid is closest to the query point and interpolate with respect to that triangle. For reasonably smooth input meshes with roughly uniform triangle sizes, the resulting points lie very close to the original surface, and for the vertices themselves the mapping stays exact. For sufficiently dense input meshes an even simpler variant is possible: assign to each query point the 3D position of the nearest embedded vertex. This has the advantage that all mapped points lie exactly on the input surface instead of merely close to it, at the cost of a piecewise-constant (i.e. non-smooth) mapping.



(a) (b) (c) (d) (e)

Fig. 3: Embedding a surface patch into a rectangular region: (a) input mesh (b) embedded point cloud (c) transferred mesh data (d) relocated boundary (e) relaxed interior

Using this extended mapping, we place a regular two-dimensional grid in the parameter region (Figure 4a) and map it into 3D. The resulting 3D point grid is then used directly as the control point grid of a B-Spline surface (Figure 4b). This B-Spline surface (optionally slightly extended beyond the data boundary) serves as the initial guess for the fitting routine of the *ON_nurbs* module of PCL [12] (Figure 4c). The routine first estimates, for every input point, a corresponding parameter position on the given surface and then refines these positions iteratively. This works reliably for a flat initial surface obtained from a PCA fit, but for our generally curved initial surface the automatic estimate is not always correct. We therefore hand the already known parameter positions of the input vertices to the fitting routine as well. With a suitable choice of fitting parameters, the procedure yields a B-Spline surface that closely approximates the original input data (Figure 4d). The accuracy of the resulting surface can be adjusted by increasing approximation thresholds and allowing further knot insertion. Theoretically, the procedure allows for arbitrarily precise approximation, but in practice it is a trade-off between accuracy and computational time. The computational effort caused by knot insertions grows exponentially.

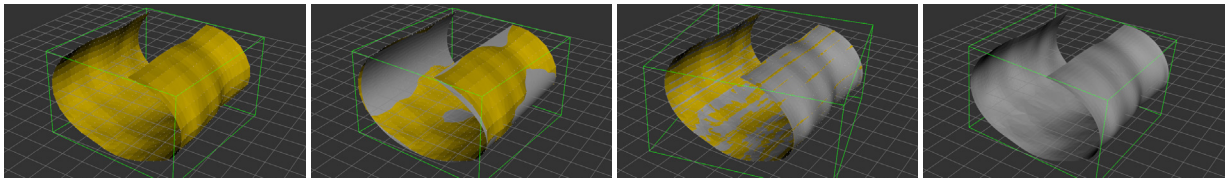


(a) (b) (c) (d)

Fig. 4: (a) grid in parameter region (b) B-Spline surface using mapped grid as control points (c) B-Spline surface and the initial mesh data, (d) slightly extended final approximation using the fitting algorithm provided by PCL

4 Validation

We present a validation for a) the surface approximation, using the curved surface patch presented in Section 2, and b) the overall process for difference volume extraction as described in Section 3, using a damaged gear cutting tool. **a) Surface approximation validation:** Applying our approach for the initial embedding and surface, the reconstruction of the curved surface patch is successful (Figure 5). No folding occurs in the subsequent iterative fitting process, due to correct initial parametrization of the scan data through the 2D embedding.



(a) (b) (c) (d)

Fig. 5: Successful process (a) dataset (b) dataset and initial surface (c) dataset and approximation (d) approximation

b) Difference volume reconstruction validation: The complete process chain was tested with a gear cutting tool, with a diameter of $120\,\mathrm{mm}$ that exhibits a broken tooth. The defective area is an arbitrarily shaped surface. Figure 6 shows the intermediate results of the relevant process steps: captured 3D data, registration, difference segments, defect segment, defect approximation and final difference volume. Note that the defect surface was extended properly, preventing intersection errors during CAD trimming. To assess accuracy, a deviation analysis based on the trimmed

CAD and as-is scan data was carried out (Figure 6j). Only one additional knot insertion and a loose approximation threshold were used, in order to achieve a comparatively short computational time of approximately $4\,\text{min}$. The plot shows that most areas of the surface lie within $\pm 0.05\,\text{mm}$ of the scan data, which is sufficient for the intended additive repair processes. On a very limited section with high local curvature the deviation slightly exceeds $\pm 0.1\,\text{mm}$. If this were rated outside the repair tolerance, process parameters could be refined to reach the desired accuracy.

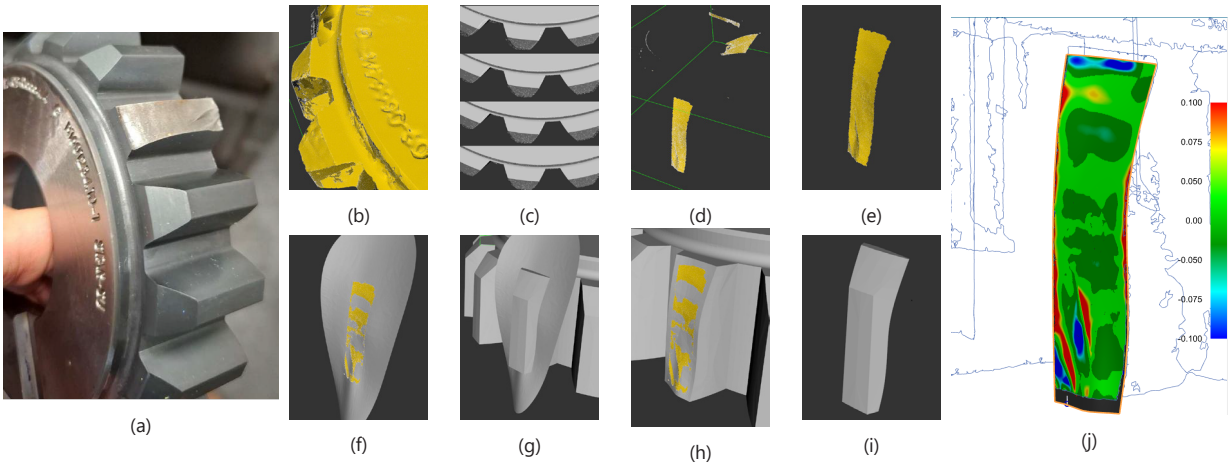


Fig. 6: Validation with gear cutting tool (a) defective part (b) captured scan data (c) multi-level registration (d) difference surfaces (e) filtered defect segment (f) segment and approximation (g) approximation and as-designed CAD (h) trimmed CAD and segment (i) final difference volume after CAD trimming (j) deviation analysis

5 Discussion and Conclusion

This paper presented an automated process chain for reconstructing arbitrary 3D defect geometries and their corresponding difference volumes from 3D scan data, intended as a building block for adaptive repair processes in MRO. The pipeline combines pre-processing, multi-level registration, correspondence-based segmentation and filtering of defect areas, and a NURBS approximation of the defective surface, which trims the as-designed CAD model. The result is a parametrized as-is geometry together with the difference volume to be restored, i.e. the input required for CAM path planning and additive repair. The results indicate that the decisive factor for a robust NURBS fit is not the fitting routine itself but the quality of its initialization. Replacing the projection onto a base surface by a geodesic-distance-approximating embedding removes the ambiguity that causes folds and self-intersections on strongly curved patches, and supplying the known parameter values of the input points avoids the misassignments that a curved initial surface would otherwise provoke. Because the embedding is derived from k-nearest-neighbour relations only, mesh connectivity is convenient but not mandatory: with the nearest-vertex variant of the mapping the procedure also applies to unstructured point clouds, which makes it usable directly on scan data. Several limitations have to be noted: First, the achievable accuracy is a trade-off against computational effort: arbitrary precision is theoretically possible, but the effort caused by knot insertion grows exponentially and is not yet quantified in terms of runtime and target tolerance. Second, the benefit over projection-based initialization is shown qualitatively but not quantitatively, and both the approximation and the process chain are demonstrated on one example each, leaving the transferability to further defect classes open. Third, the accuracy of the difference volume is bounded by the registration, which may itself be biased by the defect, as the missing material enters the alignment as systematic outliers; larger defects therefore require outlier-rejecting registration and a dedicated sensitivity analysis. Overall, the proposed process proves to be a versatile and robust approach for the automated generation of difference volumes from arbitrary macroscopic defects. Future work will address optimizing the trade-off between computational effort and accuracy, patches with non-trivial topology, as well as the integration into a fully adaptive repair process for large parts, where only a limited area around the defect is captured.

References

[1] S. Nandi, M. Mukherjee, and S. K. Das, "Digital twin-empowered sustainable remanufacturing of high-value components: a review of intelligent repair strategies," *Computers & Industrial Engineering*, vol. 220, 2026.

[2] F. Hou, N. Wan, Z. Chang, Z. Chen, and H. Sun, "An adaptive repair surface modeling approach for worn blades," *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 1-4, pp. 523–532, 2018.

[3] I. Alfred, M. Nicolaus, J. Hermsdorf, S. Kaierle, K. Möhwald, H.-J. Maier, and V. Wesling, "Advanced high pressure turbine blade repair technologies," *Procedia CIRP*, vol. 74, pp. 214–217, 2018.

[4] J. Kuschan, V. Müller, S. Monchinger, O. Heimann, C. Niebuhr, and O. Kabha, "Production environment of tomorrow (promo): Partially automated repair process of small tool moulds, forming tools, injection moulding tools and sand casting tools," *IOP Conference Series: Materials Science and Engineering*, vol. 1140, no. 1, p. 012036, 2021.

[5] "Din en 13306:2018-02, instandhaltung_- begriffe der instandhaltung; dreisprachige fassung en_13306:2017."

[6] M. J. Kim, M. Praniewicz, T. R. Kurfess, and C. Saldana, "Adaptive repair and digitization for hybrid manufacturing," *Procedia Manufacturing*, vol. 34, pp. 154–160, 2019.

[7] R. Joost, S. Mönchinger, L. Engeländer, and K. Lindow, "Automated 2d-3d-mapping and assessment of defects obtained from 2d image detection on a 3d model for efficient repair of industrial turbine blades," in *Sustainable Design and Manufacturing 2023* (S. G. Scholz, R. J. Howlett, and R. Setchi, eds.), vol. 377 of *Smart Innovation, Systems and Technologies*, pp. 79–89, Singapore: Springer Nature Singapore, 2024.

[8] R. Joost, B. Hein, B. Neumann, D. Vöhringer, H. Lammers, and K. Lindow, "Automated scan-based remanufacturing process chain for high-precision milling of turbine blades," in *Sustainable Design and Manufacturing 2025* (M. Jolly, S. G. Scholz, R. J. Howlett, and R. Setchi, eds.), vol. 483 of *Smart Innovation, Systems and Technologies*, pp. 73–82, Cham: Springer Nature Switzerland and Imprint Springer, 2026.

[9] S.-L. Nguyen, O. Ahmed, R. Joost, P. Lünnemann, and T. Larusch, "Digital twin for automated post-processing chain in additive manufacturing," in *Flexible automation and intelligent manufacturing: the future of automation and manufacturing: intelligence, agility, and sustainability* (K. Srihari, M. T. Khasawneh, S. Yoon, and D. Won, eds.), Lecture Notes in Mechanical Engineering, pp. 119–127, Cham: Springer, 2026.

[10] L. Piegl and W. Tiller, *The NURBS Book*. Springer-Verlag Berlin Heidelberg GmbH, second edition ed., 1997.

[11] Z. Zhu, A. Iglesias, L. You, and J. J. Zhang, "A review of 3d point clouds parameterization methods," in *Computational Science – ICCS 2022* (D. Groen, C. de Mulatier, M. Paszynski, V. V. Krzhizhanovskaya, J. J. Dongarra, and P. M. A. Sloot, eds.), vol. 13352 of *Lecture Notes in Computer Science*, pp. 690–703, Cham: Springer International Publishing, 2022.

[12] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *IEEE International Conference on Robotics and Automation (ICRA)*, (Shanghai, China), IEEE, May 9-13 2011.

[13] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[15] W. Schroeder, K. Martin, and B. Lorensen, *The visualization toolkit: An object-oriented approach to 3D graphics ; visualize data in 3D - medical, engineering or scientific ; build your own applications with C++, Tcl, Java or Python ; includes source code for VTK (supports Unix, Windows and Mac)*. Clifton Park, NY: Kitware Inc, 4. ed. ed., 2006.

Evaluation of an AI-Based Concept for 2D-to-3D Defect Detection in Industrial Selective Repair Processes

Dina Mahdi-Joest[1], Jolie Nimz[1], Martin Pape[1], Tom Wolf[1]
[1] Fraunhofer-Institut für Produktionsanlagen und Konstruktionstechnik IPK,
Pascalstraße 8-9, 10587 Berlin, tom.wolf@ipk.fraunhofer.de, www.ipk.fraunhofer.de

Abstract: Acquiring a complete high-resolution 3D representation of an industrial object for defect inspection is computationally demanding, often impractical under processing-time constraints and, depending on the technology, blind to non-volumetric anomalies. A more efficient strategy is to screen globally for defects in high-resolution 2D images and localize only the detected defect regions in 3D, enabling selective local surface reconstruction - the foundation of local repair strategies, in which the measured defect volume is refilled additively. We present a modular pipeline that follows this strategy, combining an RGB-D camera, unsupervised anomaly detection, a zero-shot 6-DoF pose estimation and ray casting to project each detected defect pixel onto the CAD mesh. Rather than evaluatingsimulation the full pipeline, two experiments isolate its 2D-to-3D localization and quantify its accuracy under computational constraints. Because the same defect surface is hit in many consecutive frames, naively accumulating every projected coordinate over time lets the data grow without bound. The first experiment compares a tessellation-dependent and a distance-based aggregation algorithm with regard to localization accuracy and runtime efficiency. For real industrial scenarios the error propagated over the the 2D-to-3D projection chain is the essential criterion, so each component's error is modeled separately. In the second experiment a Monte-Carlo simulation evaluates the interaction between the resolution-related errors, the pose estimation error and aggregation inaccuracies for multiple geometries, referenced against the theoretical resolution limit, the smallest resolvable distance per pixel. The proposed method evaluates the accuracy of interchangeable components in a 2D-to-3D projection meta pipeline under controlled conditions. Two approaches for temporal defect aggregation are compared based on their respective constraints. This enables a realistic estimation of defect localization accuracy, providing a basis for efficient local repair strategies.

Keywords: 2D-to-3D mapping; industrial visual inspection; defect localization; raycasting; additive repair; anomaly detection

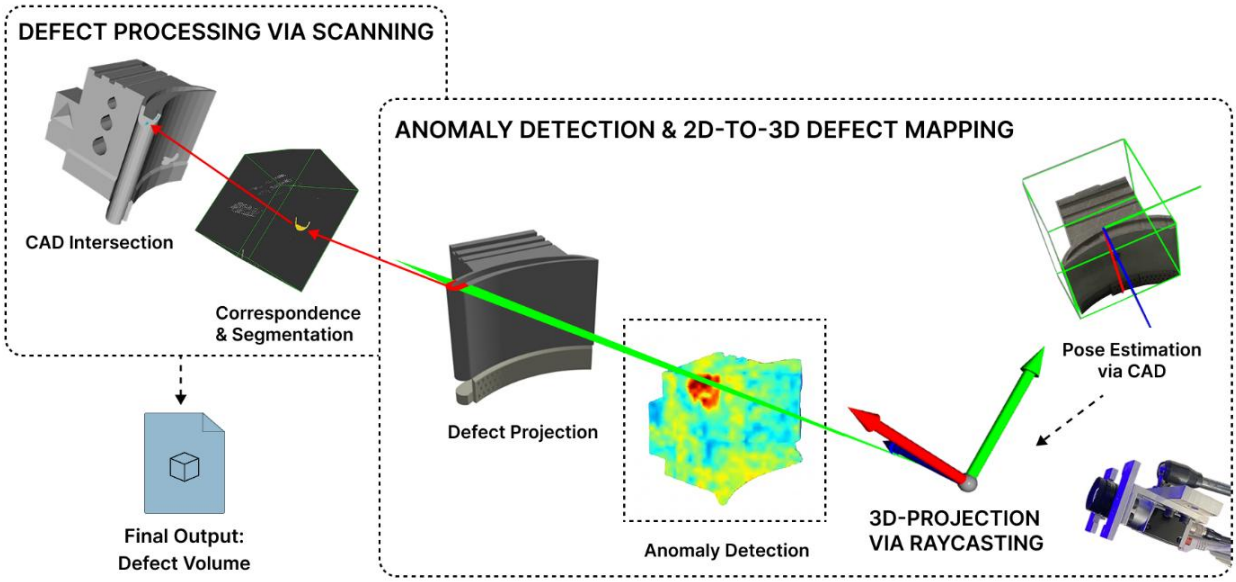


Fig. 1 Overview of the inspection concept. A 2D defect mask from anomaly detection is back-projected onto the CAD model by ray casting using the estimated 6-DoF pose (right box, the geometric 2D-to-3D transfer evaluated in this work); the localized region then drives downstream defect processing via targeted scanning, yielding the final defect volume for selective repair (left box, downstream context).

1 Topic

Local repair of defective areas is a sustainable alternative to replacing entire components. For volumetric defects, localization is typically performed by a 3D scan of the complete part. This procedure is time-consuming and may fail to detect shallow surface defects. The defects addressed in this work are surface damages that develop during the life of heavy forming tools, such as stamping and forming dies, including wear marks, scratches, spalling and corrosion pits, with lateral extents from a few millimeters to a few centimeters. Although smaller in depth, these surface-level defects still represent volumetric material loss that requires precise 3D localization for selective repair. Therefore, this work proposes a global 2D defect detection approach based on anomaly detection (AD) in combination with a 6-DoF pose estimation model. By applying ray casting, defect points identified through fast 2D detection are mapped onto the 3D model of the component, enabling their spatial localization for selective scanning and further repair processes (*Fig. 1*). An arbitrary defect detection model is assumed, generating a pixel-accurate mask of the defect region. The focus however lies in the geometric 2D-to-3D transfer and its trustworthiness, evaluated with two simulation studies that reuse the same projection code used in the actual system. Because runtime is critical in practice, the projection accuracy is studied as a function of runtime-relevant parameters.

2 Scientific Background and Motivation

Manual inspection and reconditioning of heavy tools is slow, operator dependent and error prone. Modern additive manufacturing methods enable local spot repair, in which a defined volume is refilled with new material instead of replacing the part, which is favorable for a sustainable lifecycle improvement of high-value production tools. An automated system must report not only whether a part is defective but where the damage lies in 3D in order to accurately measure the missing volume and define the desired shape. This volume follows from comparing a high-resolution local surface reconstruction with the nominal CAD geometry. Because defective examples are rare in production, detection is posed as unsupervised anomaly detection trained only on defect-free images [1], [2]. However, image-based detection is only the first step: for repair, the defect must be localized accurately in the 3D geometry of the component, yet the accuracy of this mapping is rarely quantified. Two questions govern it: how close the achievable accuracy comes to the theoretical spatial resolution limit imposed by the camera, and how repeated detections of the same defect should be consolidated into an efficient, non-redundant 3D representation. The distinction from existing 2D-to-3D defect localization approaches is therefore not a new projection method, but a component-wise error budget: detector, pose estimator and aggregation criterion can be exchanged, while the resulting chain accuracy is quantified rather than assumed. Practical advantages include shorter capture times due to local 3D scanning of only those defect areas that have been globally detected in 2D, sensitivity to surface defects, and - with regard to anomaly detection methods - training based exclusively on defect-free images.

3 Methods

Pipeline. The evaluated pipeline integrates image-based defect detection, model-based pose estimation, and geometric projection to obtain 3D defect locations for subsequent selective repair. *Fig. 1* illustrates the complete processing pipeline. First, RGB-D frames are acquired with a stereo depth camera and processed by object segmentation, zero-shot 6-DoF pose estimation, unsupervised anomaly detection, and ray-casting-based 2D-to-3D projection. In the deployed system, anomaly detection uses Dinomaly [3], an encoder-decoder reconstruction approach on frozen DINOv2 features [4], although the concept evaluation is agnostic to the specific detector. As with most feature-map-based architectures, its resulting defect map shows a reduced resolution compared to the input, which motivates treating the mask resolution as an explicit error source below. The 6DoF-pose is estimated by FoundationPose [5], a model-based zero-shot estimator that generalizes to novel objects without retraining and estimates the rotational and translational components of the object pose in camera coordinates from an RGB-D frame. For each pixel of the binary defect mask, a ray is cast onto the component surface to obtain the corresponding 3D defect location. Whether achieved by moving the object relative to the camera or by using multiple viewpoints, repeated observations of the same defect produce 3D projections that are accumulated over time. The localized defect region then serves as input for high-resolution local scanning and subsequent material deposition. Two simulation experiments evaluate the properties of the projection systematically.

Experiment 1: Defect Accumulation. When the same defect is hit in several consecutive frames, the projected 3D points must be reduced into one consistent 3D representation instead of being accumulated repeatedly. Two aggregation algorithms are compared (*Fig. 2*): *triangle-ID binning*, which keeps only the first hit per mesh triangle, and a *radius-neighborhood* criterion that discards new points within a radius *r* of an already aggregated point. We refer to this criterion as "physically controlled" because r is specified in millimeters independently of mesh tessellation. Three geometries, a sphere, a cuboid and a geometrically complex real component, carry defect regions defined independently of the mesh triangulation, so that neither algorithm is structurally favored. The three defect types are modeled as a circular patch (Ø 20 mm), a rectangular patch (20×12 mm) and an elongated stripe (7×40 mm). Each defect type is additionally evaluated at half and double size. The single-view condition uses one frontal frame. To assess the effect of repeated observations, the multi-view condition adds eight randomized poses drawn uniformly around the object, giving nine overlapping views of which seven show the defect. All three geometries share this pose set. Each algorithm is evaluated against a defect region by the reconstruction overlap (IoU), the mean coverage error (mean distance from the true defect surface to the nearest kept point), the point-density uniformity (coefficient of variation (CV) of nearest-neighbor spacing), and the deduplication runtime. The radius *r* is swept from 0.1 to 1.6 mm in 0.1 mm steps. Tessellation is varied over six subdivision levels for the sphere and the cuboid, and over three levels of the real component. Every configuration is evaluated with five seeds generating different ground-truth samplings to separate real effects from measurement noise.

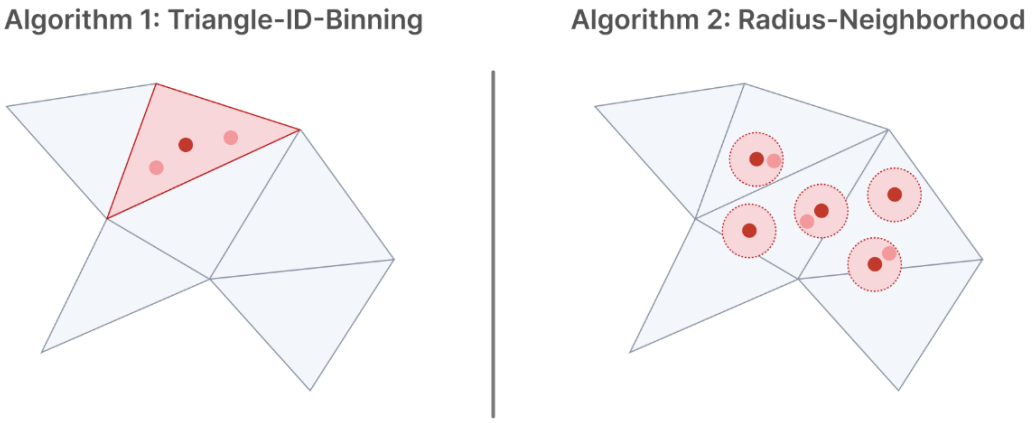


Fig. 2 Defect aggregation strategies: *Triangle-ID binning keeps one point per mesh triangle; the radius criterion keeps points spaced by r, independent of the tessellation. Dark red points are kept, faded points are discarded.*

Experiment 2: Error-Chain Analysis. A Monte-Carlo simulation is used to quantify the localization error of a single processed frame, modeled through four contributions (*Fig. 3*): the defect-mask resolution (E1), ray-grid resolution (E2), pose estimation error (E3), and defect-accumulation error (E4). E1 is emulated by down- and re-upsampling the ground-truth mask by factors 1-36, corresponding to mask edge lengths from 1024 px down to 28 px, where the coarse end matches the feature-map resolution of real detectors [3]. E2 uses an n×n ray grid with n from 2048 to 28 px. Both are hyperparameters of the simulation and are swept jointly. E3 models the pose estimation error by perturbing the ground-truth pose with random offsets drawn from the empirical error distribution of FoundationPose [5], obtained by running it in tracking mode over the YCB-Video test split. The 14025 keyframe poses were validated against the published benchmark scores (AUC of ADD-S 97.5 vs. 97.9 and ADD 96.0 vs. 96.0 after excluding one rotationally symmetric object), confirming correct units, coordinate conventions, and symmetry handling. The per-axis error distributions were extracted for nearly unoccluded views with a visible fraction above 0.9, matching the essentially unoccluded presentation in the industrial cell, and yield lateral and axial translation sigmas of 1.94 and 1.64 mm. The model draws independent zero-mean Gaussian errors per axis, with rotation sigmas of 1.98° for tilt and 1.25° about the optical axis, plus the measured constant bias of +1.0 mm along the optical axis, all applied in the camera frame. Axis independence and the Gaussian core were verified on the data. On the simulated objects this noise reproduces the pose quality level of the public benchmark in terms of the standard ADD metric [5]. E4 is a constant offset from Experiment 1 for the matching geometry and defect type at the operating point. The simulation spans all combinations of 3 geometries, 3 defect types, camera distances Z ∈ {200, 300, 500,

700, 900, 1100} mm, incidence angles θ ∈ {0°, 30°, 60°}, 7 E1 and 8 E2 levels. Each combination is evaluated with 15000 pose draws, sufficient for a statistical uncertainty of the cell means below ±1.5 %, and computed twice, once with the measured pose noise and once with E3 set to zero. The control run yields the residual error of resolution and aggregation, so that the pose contribution is obtained by difference. The resulting error is related to the theoretical spatial resolution limit of the pinhole camera, expressed as the projected pixel size on the object surface

$$s = \left(\frac{Z}{f}\right) \cdot \left(\frac{1}{\cos\theta}\right) \qquad \textit{(Eq. 1)}$$

which follows from the depth-based image scale of oblique views [7] with the focal length f = 1400 px expressed in pixel units [6], i.e. 0.21 mm/px in the reference scenario at Z = 300 mm and θ = 0°. Runtime is measured in separate single-core benchmarks at identical settings.

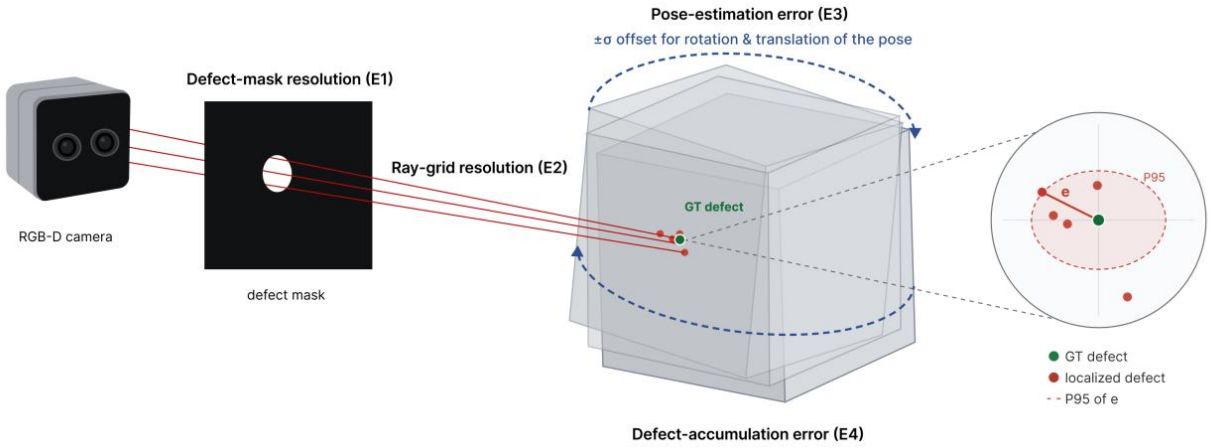


Fig. 3 Principle of the error-chain simulation. Rays are cast through the active pixels of the coarsened defect mask (defect-mask resolution, E1) on a ray grid (ray-grid resolution, E2) using a pose perturbed by the measured pose estimation error (E3), including the measured translational and rotational ±σ variations and the axial translation offset. The defect-accumulation error (E4) from Experiment 1 is then applied. The localized points (red) are compared with the ground-truth defect (green).

4 Results

Experiment 1: Defect Accumulation. compares both strategies with the no-accumulation reference.

Tab. 1 Aggregation strategies per geometry (nine poses, mean over the three defect types at Ø 20 mm, 20×12 mm and 7×40 mm and all tessellation levels; five evaluation seeds). Triangle-ID at the finest tessellation available per geometry; the radius at 0.4 mm.

Criterion	No accumulation (reference)			Triangle-ID (finest tess.)			Radius (r = 0.4 mm)		
	sphere	cuboid	real part	sphere	cuboid	real part	sphere	cuboid	real part
Points kept	30742	31117	30170	750	277	98	1437	1031	1380
IoU vs. ground truth	0.91	0.91	0.91	0.91	0.83	0.73	0.94	0.93	0.95
Mean coverage error [mm]	0.04	0.04	0.05	0.25	0.42	0.75	0.17	0.20	0.18
Point-density CV	0.46	0.46	0.51	0.11	0.07	0.25	0.04	0.01	0.02

Both reduce the number of raw projected points by more than an order of magnitude while preserving accurate coverage of the defect region. The decisive difference lies in what controls the achievable accuracy. For the radius criterion the coverage error is independent of the mesh tessellation; across all tessellation levels it varies by less than 0.01 mm. It is likewise stable across the three geometries, with an IoU of 0.93-0.95 and a coverage error of 0.17-0.20 mm at the operating point, and unaffected by halving or doubling the defect size; the achievable accuracy is set by r alone. We select r = 0.4 mm as an accuracy-oriented operating point, which reduces the point count 24-fold. A smaller radius recovers accuracy almost to the reference level, a larger radius trades it continuously for further compression. For triangle-ID binning the decisive factor is the mesh tessellation, a property of the CAD export rather than of the algorithm. The coverage error on the real component varies between 0.75 and 2.08 mm depending on

the tessellation, and even the finest available tessellation stays a factor of four above the radius criterion. Going from one pose to nine poses raises the unaggregated point count per defect from $8.5 \cdot 10^3$ to $3.1 \cdot 10^4$ and the density CV from below 0.01 to 0.46-0.51, while both strategies remain unchanged within 0.01 mm coverage error and 0.03 IoU. Both therefore bound the representation independently of the number of contributing views, and the radius criterion equalizes the density to CV ≤ 0.04. . Both algorithms process each incoming frame without reprocessing previously accumulated points, keeping the per-frame cost approximately constant. The radius criterion uses a spatial hash grid, while triangle-ID binning uses direct mesh-triangle indexing. Over 20 frames of 5000 projected points each, the per-frame runtime remained at 37.4 to 39.5 ms for the radius criterion and 0.35 to 0.47 ms for triangle-ID binning. Reducing a full set of 10^5 points in one call showed a similar difference (769 ms vs. 5.7 ms). Thus, the radius criterion incurs substantial overhead, while its advantage lies in predictable, tessellation-independent accuracy rather than computational efficiency.

Experiment 2: Error-Chain Analysis. The following values refer to the baseline scenario, a camera distance of 300 mm, at frontal view (θ = 0°) with an anomaly map of 256×256 (one cell covers 0.86 mm) and a ray grid of 1024×1024.

Tab. 2 Error-chain decomposition. Mean over the three geometries, 15000 pose draws per cell, statistical uncertainty of the means below ±1.5 %. Fine resolution settings (anomaly map = 256×256 (E1 = 4), ray grid = 1024×1024). The pose contribution is the difference to the control run without pose noise. Between 200 and 500 mm the pooled pose noise applies, above 700 mm the noise measured at that distance.

	Total error [mm]	**σ_{lat} [mm]**	**Resolution + aggregation [mm]**	**Pose term E3 [mm]**	**P95 [mm]**	**× pixel footprint**
By defect type (Z = 300 mm, θ = 0°)						
Circle	0.56	1.94	0.22	0.34	1.28	2.6
Rectangular	0.65	1.94	0.22	0.43	1.55	3.0
Stripe	0.76	1.94	0.23	0.53	2.13	3.5
By incidence angle (circular defect, Z = 300 mm)						
θ = 0°	0.56	1.94	0.22	0.34	1.28	2.6
θ = 30°	0.59	1.94	0.22	0.37	1.40	2.4
θ = 60°	0.89	1.94	0.24	0.65	2.46	2.1
By camera distance (circular defect, θ = 0°)						
200 mm	0.54	1.94	0.19	0.35	1.26	3.8
300 mm	0.56	1.94	0.22	0.34	1.28	2.6
500 mm	0.58	1.94	0.25	0.33	1.31	1.6
700 mm	0.50	1.65	0.27	0.24	1.06	1.0
900 mm	0.65	2.00	0.35	0.30	1.41	1.0
1100 mm	0.73	2.17	0.40	0.33	1.58	0.9

Each simulated frame yields the localization error, the mean distance from its reconstructed points to the true defect surface; mean and 95th percentile are taken over the 15000 frames. For the circular defect the localization error is 0.56 mm on average and stays below 1.28 mm in 95 % of the frames, corresponding to 2.6 times the projected pixel from (*Eq. 1*). The control run without pose noise leaves a residual error of 0.22 mm from resolution and aggregation alone, so the pose term contributes 0.34 mm or 61 % of the total and is the dominant source. With identical inputs the pose contribution grows to 0.43 mm (rectangular patch) and 0.53 mm (stripe), and the total error follows to 0.65 mm and 0.76 mm (Tab. 2). Within the application range from 200 to 500 mm, where the pooled noise set applies, increasing camera distance mainly enlarges the spatial sampling interval of Eq. 1 and the total error grows only from 0.54 to 0.58 mm. Above 700 mm the pose noise measured at each distance is used: the lower uncertainty at 700

mm reduces the total error to 0.50 mm before it rises to 0.73 mm at 1100 mm. This non-monotonic behavior is therefore driven by measured pose uncertainty rather than geometric projection alone. E3 accounts for approximately 46-73 % of the total error across the evaluated conditions (*Tab. 2*). Varying the two resolution parameters over more than an order of magnitude changes the mean error by less than 25%, while reducing the per-frame ray-casting cost 12 to 19× between the finest and the coarse settings. This indicates that E1 and E2 primarily provide computational trade-offs rather than dominant point-wise accuracy limitations. However, when the anomaly-map cell size approaches the defect width, coarse E1 can limit complete defect capture. For the narrow stripe, which is 7 mm wide, the coarsest 28×28 anomaly map corresponds to a cell size of 7.7 mm at Z=300 mm and therefore cannot fully resolve the defect width. Consequently, under the same resolution setting, the localization error for the stripe is substantially higher than for the circular defect, increasing from 0.39 mm to 1.56 mm.

5 Discussion

Experiment 1 demonstrates that Triangle-ID binning is computationally efficient, but its spatial resolution is implicitly determined by CAD tessellation, whereas the radius criterion sets the desired point separation explicitly. At r=0.4 mm, it reduces the representation by more than an order of magnitude while maintaining a coverage error of 0.18 mm and a uniform point density. **Experiment 2** shows that, under the baseline conditions, pose estimation is the dominant source of point-wise localization error. The defect-mask and ray-grid resolutions have a comparatively small effect on point-wise localization accuracy within the tested range, but coarse defect-mask resolutions can become limiting when the spatial resolution approaches the size of the defect itself. For selective repair, the results therefore favor a physically controlled accumulation strategy and reliable 6-DoF pose estimation, while allowing moderate reductions in detector and ray-grid resolution within this limit when runtime is critical. The study is limited by the use of programmatically defined defect regions and a controlled set of viewing and geometric conditions. Further evaluation on experimentally acquired defects and more complex surface geometries would be required to assess the robustness of the complete localization chain under industrial operating conditions.

6 Conclusion

This study evaluated an AI-based 2D-to-3D localization chain for selective repair and quantified the contributions of accumulation, resolution, and pose uncertainty to the final localization error. A physically defined radius-based accumulation criterion provides a mesh-independent representation with reduced redundancy, while pose estimation represents the dominant source of localization error under evaluated conditions. Reliable 6-DoF pose estimation and controlled accumulation proved more critical than the finest detector and ray-grid resolutions, which can be reduced moderately as long as the anomaly map still resolves the smallest defect. These findings provide a basis for selecting efficient settings for 2D-to-3D defect localization in selective repair applications.

References

[1] Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C.: MVTec AD - A comprehensive real-world dataset for unsupervised anomaly detection. In: Proc. IEEE/CVF CVPR, 2019, pp. 9592–9600. doi:10.1109/CVPR.2019.00982

[2] Roth, K.; Pemula, L.; Zepeda, J.; Schölkopf, B.; Brox, T.; Gehler, P.: Towards total recall in industrial anomaly detection. In: Proc. IEEE/CVF CVPR, 2022, pp. 14318–14328. doi:10.1109/CVPR52688.2022.01392

[3] Guo, J.; Lu, S.; Zhang, W.; Chen, F.; Li, H.; Liao, H.: Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. In: Proc. IEEE/CVF CVPR, 2025. arXiv:2405.14325

[4] Oquab, M.; Darcet, T.; Moutakanni, T. et al.: DINOv2: Learning robust visual features without supervision. Transactions on Machine Learning Research, 2024. arXiv:2304.07193

[5] Wen, B.; Yang, W.; Kautz, J.; Birchfield, S.: FoundationPose: Unified 6D pose estimation and tracking of novel objects. In: Proc. IEEE/CVF CVPR, 2024, pp. 17868–17879. arXiv:2312.08344

[6] Hartley, R.; Zisserman, A.: Multiple View Geometry in Computer Vision. 2nd ed. Cambridge: Cambridge University Press, 2004, Sec. 6.1, pp. 153–157. doi:10.1017/CBO9780511811685

[7] Maset, E.; Fusiello, A.: Scales of oblique photographs updated. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 207, 2024, pp. 198–202. doi:10.1016/j.isprsjprs.2023.12.003

Methoden zur Qualitätssicherung urbaner digitaler Zwillinge

Luna Riegel[1], Matthias Betz[1], Volker Coors[1], Ute Wagner[2], Margitta Pries[2]

[1]Hochschule für Technik Stuttgart, Fakultät Vermessung, Informatik und Mathematik
Schellingstr. 24, D-70174 Stuttgart, luna.riegel@hft-stuttgart.de
[2]Berliner Hochschule für Technik, Fachbereich II - Mathematik, Physik, Chemie
Luxemburger Str. 10, D-13353 Berlin, ute.wagner@bht-berlin.de

Abstract: This paper presents how methods for detecting and correcting errors in individual building models can be extended to urban digital twins. It focuses in particular on verifying the topological relationships among objects. A specific challenge that must be addressed is the efficient handling of massive amounts of data.

Keywords: Urban Digital Twin, Semantic 3D City Models, BIM-GIS-Integration, Topological Validation

1 Einleitung

Der Begriff des „digitalen Zwillings" für komplexe virtuelle Modelle physischer Produkte und Prozesse wurde Anfang der 2000er-Jahre in der Industrie geprägt [1]. Seitdem wird zunehmend mittels digitaler Zwillinge die Gesamtheit der Daten von Maschinen, Fahrzeugen und technischen Anlagen verwaltet, um damit deren Verhalten zu simulieren und Optimierungen vorzunehmen, bevor Veränderungen in der realen Welt umgesetzt werden. Inzwischen hat sich dieses Konzept auch für die Planung und das Management von Städten und Infrastruktur etabliert und wird in diesem Bereich stetig angepasst und weiterentwickelt. Urbane digitale Zwillinge beinhalten zusätzlich zur geometrischen Repräsentation weitere Informationen zu Eigenschaften, Funktionen und Beziehungen der Objekte des Stadt- und Landschaftsraums. Sie dienen als Datengrundlage in vielen Anwendungsgebieten wie Mobilität, Stadtplanung, Klima- und Umweltschutz für Simulationen und zur Entscheidungsunterstützung.
Für die dazu notwendige einheitliche semantisch strukturierte Beschreibung von Städten und Landschaften hat sich CityGML[1] etabliert [2], [3]. Während klassische Geoinformationssysteme auf einem 2,5D-Datenmodell — Linien und ebene Polygone oder zweidimensionale Zellraster mit Höhenangaben als Attributen — basieren, geht CityGML darüber hinaus: es beschreibt Objekte als dreidimensionale Körper über ihre Begrenzungsflächen, wobei als geometrische Primitive allerdings nur ebene Polygone zugelassen sind. Gekrümmte Formen sind somit im Modell nicht exakt abgebildet, und die Abweichung zwischen Modell und Realität ist in den CityGML-Daten nicht dokumentiert. Des Weiteren enthalten die Modelle Redundanzen und keine topologischen Informationen, woraus sich Probleme ergeben, die im Folgenden dargestellt werden.

2 Problemstellung

Während es im Bereich der Datenerfassung umfangreicher und vielfältiger, mit semantischen Informationen angereicherter 3D-Stadtmodelle bereits erhebliche Fortschritte gibt, fehlen ausgereifte Werkzeuge zur Qualitätssicherung, um eine reibungslose und fehlerfreie Verarbeitung der Daten innerhalb unterschiedlicher Prozessketten und damit eine Weiterverwendung der Modelle in vorhandenen Anwendungen im Sinne einer kontinuierlichen Integration, wie in [4] für den CAD[2]-Bereich beschrieben wird, zu ermöglichen. Das System CityDoctor [5] beispielsweise stellt Methoden zur Fehlerfindung und Korrektur von einzelnen Gebäudemodellen zur

[1] CityGML: offener Standard für georeferenzierte Stadt- und Landschaftsmodelle, der vom Open Geospatial Consortium verabschiedet wurde.
[2] Computer Aided Design

Verfügung. Für urbane digitale Zwillinge sind darüber hinaus Methoden zur Prüfung der topologischen Beziehungen der Objekte untereinander erforderlich. Weiterhin benötigt eine automatisierte Korrektur der geometrischen Daten oft Entscheidungen, die ohne ein Referenzmodell dazu führen, dass die reparierte Objektgeometrie als Ganzes zwar fehlerfrei ist, aber die Realität nicht präzise genug abbildet. Zudem ist zu berücksichtigen, dass eine Integration von Daten aus dem GIS[3]- und dem BIM[4]-Bereich Probleme mit sich bringt, die unter anderem durch Unterschiede in der Modellierung, der eingesetzten Software und der offenen Standards bedingt sind [6].

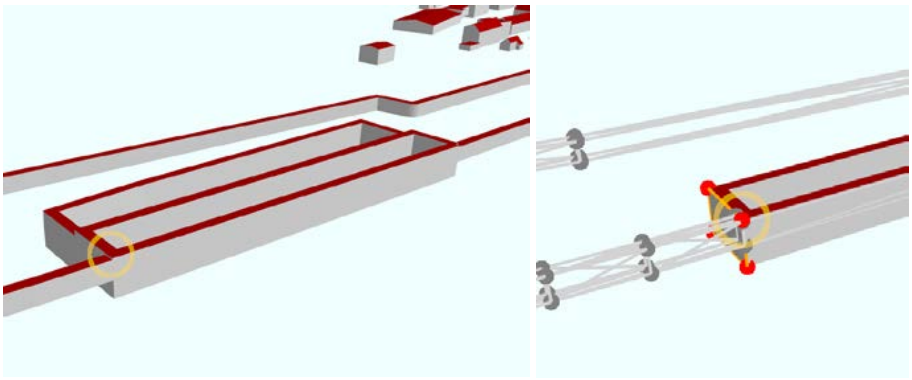


Abbildung 1: Volumetrische Überlappung von Features am Beispiel einer Schleuse

Im Folgenden wird die Prüfung topologischer Beziehungen im Rahmen des Projekts UDigiTDoctor[5] näher beleuchtet. Zum einen soll überprüft werden, ob sich die 3D-Geometrien einzelner Objekte (Features), insbesondere Gebäude, eines digitalen Zwillings volumetrisch überlappen (Featurekollision - Abb. 1). Eine entsprechende auf die Bestandteile eines einzelnen Gebäudemodells bezogene Prüfung ist im System CityDoctor bereits integriert. Für die Übertragung bzw. Weiterentwicklung dieser Prüfung auf urbane digitale Zwillinge wurde ein internes Datenmodells entworfen, um in unterschiedlichen Formaten vorliegende und teilweise aus unterschiedlichen Referenzmodellen stammende Modelldaten integrieren zu können. Eine besondere Herausforderung ist hierbei das Datenmanagement. Urbane digitale Zwillinge können sehr hohe Detailgenauigkeit besitzen. Zudem kann das Modell durch die Integration zusätzlicher Datensätze wie beispielsweise unterirdischer Ver- und Entsorgungsleitungsnetzwerke mehrschichtig sein, wodurch die Dateigrößen recht schnell ansteigen. Dies ist insbesondere problematisch, da die Daten nach dem Einlesen aufgrund des Java Object-Overheads signifikant mehr Speicher benötigen. Tabelle 1 demonstriert diesen Mehrverbrauch anhand eines Beispiels (Abb. 2).

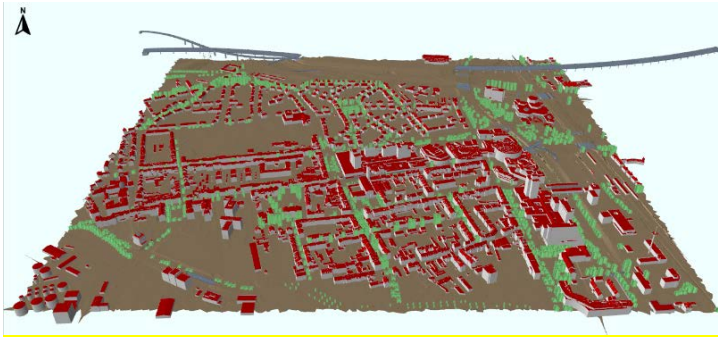


Abbildung 2: Beispiel einer CityGML Kachel, Ausschnitt aus einem digitalen Stadtmodell von Wiesdorf (Leverkusen)

Name	Wiesdorf
Kachelgröße	2x2 km
Dateigröße	170 MB
Anzahl Objekte	~ 3700
Arbeitsspeicherverbrauch nach Einlesen	~ 1 GB

Tabelle 1: Metadaten der Wiesdorf CityGML-Kachel

[3] Geographische Informationssysteme
[4] Building Information Modeling

Das Problem des Datenmanagement wird durch die räumlich-topologischen Prüfungen weiterhin verstärkt, da diese, im Gegensatz zu den geometrischen und semantischen Prüfungen, Querzugriffe auf andere Features im Modell benötigen. Aufgrund der Dateigrößen ist es jedoch oft nicht möglich, das komplette Modell im Arbeitsspeicher zu halten. Die bisherig genutzte Methodik des sequentiellen Einlesens, Validierens und anschließenden Verwerfens der Java-Objekte der Features ist jedoch nicht für die räumlich-topologischen Prüfungen geeignet. Dies erzwingt die Implementierung eines Systems, welches die Modelldaten außerhalb des Arbeitsspeichers hält, aber dennoch schnellen Zugriff auf die Feature-Daten ermöglicht.
Des Weiteren skaliert die Anzahl an Feature-Paaren quadratisch mit der Anzahl an Features. Hierdurch wird ein Prüfen aller möglichen Feature-Paare ein unbewältigbarer Rechenaufwand, was zudem durch den größeren Rechenaufwand der räumlich-topologische Prüfungen verstärkt wird. Jedoch ist es nicht notwendig, alle Feature-Paare in einem Modell zu vergleichen, da z. B. eine Featurekollision zwischen Features, deren Lage sich stark unterscheidet, direkt ausgeschlossen werden kann. Folglich können unkomplizierte und schnelle Berechnungen genutzt werden, um die Features nach „Kandidaten" zu filtern, für welche eine genauere, aufwändigere Berechnung gerechtfertigt ist.

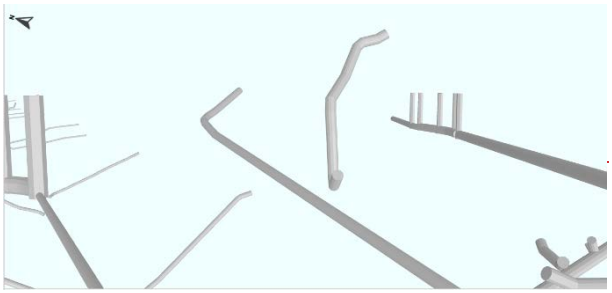


Abbildung 3: Beispiel eines Leitungsnetzmodells mit freischwebenden Leitungsabschnitten

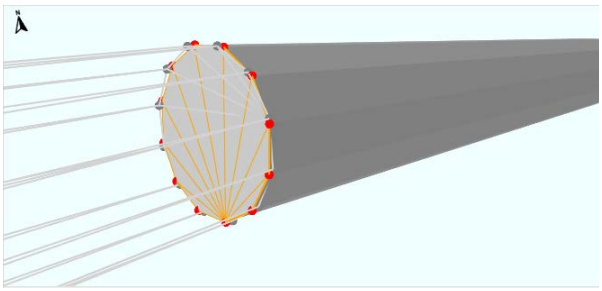


Abbildung 4: Abweichung der Knoten der ClosureSurfaces zweier Leitungsobjekte

Ein weiteres Aufgabengebiet betrifft die Überprüfung der topologischen Konsistenz urbaner digitaler Zwillinge für Leitungsnetze. Der Prüfungs- und Reparaturumfang für den Anwendungsfall Leitungsnetze wurde im Rahmen des Projekts zunächst in Kooperation mit Partnerfirmen erarbeitet. Neben dem großen zu prüfenden Datenvolumen ergeben sich zusätzliche im Format CityGML begründete Herausforderungen z.B. bei der Überprüfung von Anschlüssen der einzelnen Komponenten eines Leitungsnetzes (Abb. 3, 4).

3 Methodik

Um das Datenmanagementproblem zu lösen, wurde eine räumliche Datenbank in das Datenmodell integriert, welche als Zwischenspeicher der Features des Stadtmodells dient. Sie ermöglicht die für die räumlich-topologische Validierung benötigten Querzugriffe im Modell, ohne dass das gesamte Stadtmodell im Arbeitsspeicher gehalten werden muss. Allerdings verlängert sich dadurch die Zugriffszeit auf einzelne Objekte deutlich, da diese zunächst Deserialisiert werden müssen. Weil während der räumlich-topologischen Prüfung eines Features vermehrt auf dieselben Features in relativer Nähe zugegriffen wird, wurde zusätzlich ein Cache implementiert, welcher einen Teil des digitalen Zwillings im Arbeitsspeicher behält und somit wesentlich schnelleren Zugriff auf diese Objekte ermöglicht.
Die räumliche Datenbank erlaubt es zudem, räumliche Abfragen für die Auswahl von Kandidaten zu nutzen: Es wird überprüft, ob sich die zweidimensionalen Axis-Aligned Bounding Boxes (AABB) zweier Features überschneiden. Ist dies der Fall, so sind die Features ein Kandidatenpaar.

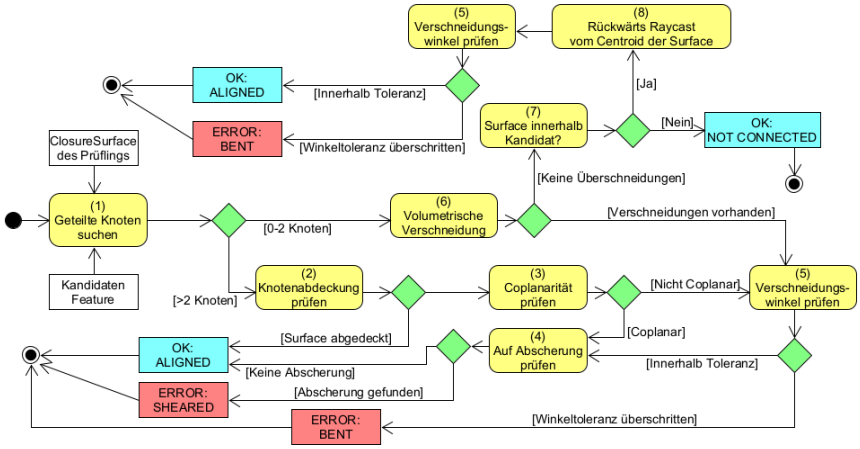


Abbildung 5: Vereinfachter Ablauf der Leitungsanschlussprüfung

Ein Beispiel für eine räumlich-topologische Prüfung aus der Leitungsplanung ist die Prüfung der Leitungsanschlüsse und Verbindungswinkel. Sie stellt fest, ob die Elemente eines Leitungsnetzes korrekt miteinander verbunden sind und ob die Verbindungen in einem zulässigen Winkel vorliegen. Zudem soll geprüft werden, ob einzelne Abschnitte vom Gesamtnetz getrennt sind (siehe Abb. 3).

Eine Herausforderung dieser Prüfung ist, dass das konzeptuelle Modell von CityGML keine eigenen Klassen für Leitungsobjekte vorsieht. Leitungsobjekte müssen daher als Gebäude oder allgemeine Bauwerke abgebildet werden, sodass sie sich nicht direkt über ihren Objekttyp abfragen lassen. Ver- und Entsorgungsleitungen sind funktionsbedingt Hohlkörper. Werden sie entsprechend modelliert, sieht CityGML virtuelle Verschlussflächen („ClosureSurfaces") vor, welche die geometrische Hülle des Features schließen. Das Vorhandensein solcher Verschlussflächen wird daher in der momentanen Version der Prüfung genutzt, um Leitungsobjekte zu identifizieren. Da CityGML-Geometrien aus Polygonen aufgebaut sind, und keine Unterstützung für Kurvenlinien bzw. Splines kann die Form von Leitungsobjekten nur annäherungsweise abgebildet werden. Die Anschlussprüfung nutzt dies, um in einem mehrstufigen Verfahren (siehe Abb. 5, folgend nur 5.) die Verbindungen zu prüfen. Zunächst wird geprüft, ob die ClosureSurfaces zweier Leitungsobjekte deckungsgleiche Knotenpositionen (geteilte Knoten) besitzen (5.1). Ist dies der Fall, wird überprüft ob alle Knoten der ClosureSurface des Prüflings geteilte Knoten sind (5.2). Ist dies der Fall, so sind die Leitungsenden verbunden. Andernfalls wird mit Hilfe der geteilten Knoten eine Verschneidungsebene berechnet, welche zum Prüfen der Koplanarität genutzt wird (5.3). Liegen alle Knoten der ClosureSurface auf der Ebene oder innerhalb einer vorherig festgelegten Toleranz (Standardwert: 1 cm), besteht möglicherweise eine Verbindung zwischen den Leitungsobjekten. Zur Kontrolle werden die Mittelpunkte der ClosureSurface Geometrien (Centroide) von Prüfling und Kandidat berechnet und deren Abstand zueinander bestimmt. Ist der Abstand innerhalb der vordefinierten Toleranz, so liegt eine Verbindung vor. Andernfalls ist die Verbindung „abgeschert" und wird mit einem entsprechenden Fehler markiert (5.4).

Liegt keine Koplanarität vor, wird die Verschneidungsebene genutzt, um die an der Verschneidung beteiligten Polygone der ClosureSurfaces beider Leitungsobjekte auszuwählen und den lokalen Verschneidungswinkel zu berechnen (5.5). Für jedes Polygonpaar (P_a, P_b) wird mit dem Möller-Trumbore-Verfahren [7] geprüft, ob sich diese schneiden, sowie die Länge l_i der Verschneidung. Anschließen wird der Diederwinkel zwischen den Flächennormalen (n_a, n_b) dieses Paares berechnet:

$$\theta_i = atan2(||n_a \times n_b||, n_a \cdot n_b)$$

Bei einem idealen, geradlinigen Anschluss zeigen die Flächennormale der ClosureSurfaces in entgegengesetzte Richtungen. Folglich wird die Abweichung gegenüber einer geradlinigen Verbindung bestimmt:

$$\beta_i = 180° - \theta_i$$

Die Winkelabweichungen aller Polygonpaare werden zu einem mit der Schnittlänge gewichteten Mittelwert zusammengefasst:

$$\beta = \frac{\sum \beta_i * l_i}{\sum l_i}$$

Anschließend wird β mit der vorherig festgelegten Winkeltoleranz (Standardwert: 3°) verglichen. Liegt β innerhalb der Toleranz, folgt die Prüfung auf Abscherung (5.4). Andernfalls wird die Verbindung als abgeknickt bewertet und mit dem Fehler „BENT" markiert.
Liegen höchstens zwei geteilte Knoten zwischen den ClosureSurfaces vor (0–2 Knoten), lässt sich aus den Knoten keine Verschneidungsebene bestimmen, da hierfür mindestens drei Punkte erforderlich sind. Stattdessen wird die volumetrische Verschneidung der ClosureSurface mit der kompletten Geometrie des Kandidaten berechnet (5.6). Liegen Verschneidungen vor, wird der Verschneidungswinkel wie in (5.5) bestimmt.
Andernfalls wird geprüft, ob die Verschlussfläche vollständig innerhalb des Kandidaten liegt (5.7). Dazu wird vom Centroid der ClosureSurface ein Strahl in Richtung der Flächennormalen ausgesendet und die Anzahl der Schnitte mit der Hülle des Kandidaten gezählt. Bei einer ungeraden an Anzahl an Schnitten liegt der Centroid innerhalb des Kandidaten, andernfalls liegt er außerhalb und es besteht keine Verbindung. Dies ist vorerst kein Fehler. Ein freischwebender Leitungsabschnitt wird erst gemeldet, wenn für keinen der Kandidaten eine Verbindung gefunden werden kann (Abb. 3).
Liegt der Centroid innerhalb des Kandidaten, so ist auch die ClosureSurface komplett eingeschlossen. In diesem Fall wird vom Centroid ein weiterer Strahl in entgegengesetzter Richtung der Flächennormalen ausgesendet (5.8). Die Flächennormale des ersten Polygons des Kandidaten welches diesen berührt wird ausgewählt wird und zur Berechnung des Winkels zwischen den Normalen in (5.5) genutzt.

4 Evaluation

Zur Evaluation der Datenbank- und Cache-Einbindung wurde eine Reihe von Laufzeittests durchgeführt. Als Testmodell diente die Wiesdorf-Kachel (ca. 3700 Features, siehe Tabelle 1 und Abb. 2). Die Tests liefen auf einem System mit einer Intel Core i7-8850H (6 Kerne, 12 Threads, 2,6 GHz Basistakt). Gemessen wurde die Laufzeit vom Aufruf der CityDoctor-CLI bis zum Abschluss der Validierung und umfasst damit auch das Einlesen des Modells. Der Cache fasst standardmäßig 3000 Features. Zur Deaktivierung wurde er nicht entfernt, sondern auf ein Feature begrenzt. Die Parallelisierung wurde analog deaktiviert, indem die Größe des Thread-Pools von der Standardeinstellung (5 × Anzahl der Prozessoren) auf 1 gesetzt wurde.
Die Tests wurden in zwei Phasen durchgeführt. In der ersten Phase wurde nur die Geometrische Validierung durchgeführt. Zusätzlich wurde hier die letzte Version von CityDoctor vor der Implementierung der Datenbank (3.17.3) getestet, die das Modell im Streaming-Verfahren verarbeitet, um eine Referenzwert für die Laufzeiten zu haben. In Phase 2 wurde zusätzlich die Prüfung auf Featurekollision aktiviert. Die Leitungsanschlussprüfung wurde nicht getestet, da das Wiesdorf-Modell keine Leitungsnetze enthält.

Version	3.17.3	3.18.3			
Methodik	Sequentiell (Stream)	Parallel, mit Cache	Parallel, ohne Cache	Sequentiell, mit Cache	Sequentiell, ohne Cache
Zeit (s)	36	52	54	170	464

Tabelle 2: Ergebnisse aus Testphase 1

Methodik	Parallel, mit Cache	Parallel, ohne Cache	Sequentiell, mit Cache	Sequentiell, ohne Cache
Gesamtzeit (s)	101	158	662	1343
Davon Featurekollision (s)[6]	49	104	492	879

Tabelle 3: Ergebnisse aus Testphase 2

Wie erwartet verlangsamt die Datenbank die Verarbeitung. Sequentiell benötigt Version 3.18.3 mit Cache 170 s statt 36 s (Faktor 4,7) und ohne Cache 464 s (Faktor 12,9). Die Parallelisierung gleicht dies weitgehend aus, parallel mit

[6] Gesamtzeit abzüglich der Zeit aus Phase 1 (gleiche Methodik)

Cache lag die Laufzeit mit 52 s nur noch beim 1,4-Fachen von Version 3.17.3. Der Cache beschleunigt die sequentielle geometrische Validierung deutlich (464 s auf 170 s, Faktor 2,7), hat im parallelen Durchlauf aber kaum Wirkung (54 s gegenüber 52 s). Dieser Unterschied wird durch das Einlesen des Modells verursacht, da der Cache hierbei befüllt wird. Bei der Featurekollisionsprüfung, welche Querzugriffe auf Nachbarfeatures benötigt, beschleunigt der Cache dagegen auch den parallelen Durchlauf deutlich. Die Laufzeit der Kollisionsprüfung sinkt von 104 s auf 49 s (Faktor 2,1), sequentiell von 879 s auf 492 s (Faktor 1,8). Zu beachten ist, dass die Wiesdorf-Kachel mit rund 3700 Features nur wenig größer ist als der Cache (3000 Features). Ein Großteil des Modells passt daher in den Cache. Bei größeren Modellen ist zu erwarten, dass der Vorteil des Cache geringer ausfällt.

5 Ausblick

Die Auswahl der Kandidaten basiert momentan auf den zweidimensionalen Bounding Boxes der Features. Dies ist eine technische Limitierung durch die räumliche Datenbank, da diese nur räumliche Indizes für zweidimensionale Geometrien unterstützt. Da digitale Zwillinge jedoch eine Mehrschichtigkeit vorweisen können, werden hierdurch Features als Kandidaten ausgewählt, welche aufgrund ihres Höhenunterschieds sich nicht tatsächlich in relativer Nähe zueinander befinden. Aus diesem Grund ist die Integration eines Bounding Volume Hierarchy Trees momentan in Bearbeitung, welcher eine schnelle Filterung der Kandidaten mit Berücksichtigung der Höheninformation erlaubt. Zudem wird angestrebt, die Nutzung von Referenzdatensätzen wie PointClouds oder 3D-Meshes für die automatisierte Korrektur der Stadtmodelle zu ermöglichen. Ein fundamentales Problem ist hierbei, dass CityDoctor keine Informationen über das reale Objekt des Features besitzt, wodurch die Reparatur nur „blind" durchgeführt werden kann. Die Referenzdatensätze sollen hierbei als Soll-Zustand dienen. Dies würde es CityDoctor ermöglichen, eine Bewertung der Ähnlichkeit zum Modell, z. B. durch die mittlere quadratische Abweichung, zu erstellen. Mithilfe der Bewertung können anschließend die Effektivität einzelner Korrekturschritte und somit verschiedene Korrekturmethoden verglichen werden. Dies würde es CityDoctor erlauben, die am besten geeignete Methode selbstständig auszuwählen, und wäre somit Grundlage für die Nutzung von maschinellem Lernen in der Korrektur.

6 Förderung

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Forschung, Technologie und Raumfahrt unter den Förderkennzeichen 13FH019KA2 und 13FH019KB2 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Literatur

[1] Grieves, M., Vickers, J. (2017): Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In: Kahlen, F.-J., Flumerfelt, S., Alves, A. (Hrsg.): Transdisciplinary Perspectives on Complex Systems. Springer, Cham, S. 85–113.

[2] Kolbe, T. H., Kutzner, T., Smyth, C. S., Nagel, C., Roensdorf, C., Heazel, C. (2021): OGC City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard. OGC Document 20-010, Open Geospatial Consortium, Wayland, MA. https://docs.ogc.org/is/20-010/20-010.html (Zugriff am 07.07.2026)

[3] Coors, V., Andrae, C., Böhm, K.-H. (2016): 3D-Stadtmodelle: Konzepte und Anwendungen mit CityGML. Wichmann, Berlin/Offenbach.

[4] Frazelle, J. (2021): A New Era for Mechanical CAD. Communications of the ACM, 64(10), S. 36–39.

[5] Hochschule für Technik Stuttgart: CityDoctor – Software zur Prüfung und automatisierten Reparatur virtueller Stadtmodelle. https://transfer.hft-stuttgart.de/pages/citydoctor/citydoctorhomepage/de/ (Zugriff am 07.07.2026)

[6] Technische Universität Delft: GeoBIM – Bridging the gap between Geo and BIM. https://3d.bk.tudelft.nl/projects/geobim/ (Zugriff am 07.07.2026)

[7]: Möller, T., Trumbore, B. (1997): Fast, Minimum Storage Ray/Triangle Intersection. Journal of Graphics Tools, 2(1), S. 21–28.

Detection of Anomalous Energy Consumption Patterns in Industrial Machines through Time Series Analysis to Improve Energy Efficiency

Deya Almanasreh[1], Sri Sai Akhila Poluri[1], Jens Kneifel[1], Robin Roj[1]
[1]Forschungsgemeinschaft Werkzeuge und Werkstoffe e. V., Papenberger Straße 49, 42859 Remscheid, Germany, kneifel@fgw.de, www.fgw.de

Abstract: Industrial machines generate large amounts of electrical time-series data that can be used to identify unusual energy-consumption patterns and load peaks. This study compares interpretable statistical methods (Global and Rolling Z-scores, Global and Rolling Modified Z-scores) with Isolation Forest and Local Outlier Factor for anomaly detection in the mean active power of industrial machines. Five machines of different types out of 73 monitored machines in two companies, each covering 163 days of ten-minute data, were used. Since labels for real anomalies are not available, synthetic short load spikes, sudden level shifts, and gradual increases were inserted into the real series at known positions. Methods were compared on separate development and holdout benchmarks using event-level and observation-level metrics, a paired measure isolating the effect of the injection, a chance analysis, and the alarm burden per machine-day. Isolation Forest (0.900) and the regularized rolling methods (0.841 to 0.875) detected most spikes, with small differences that are not interpreted as a ranking. Level shifts were detected most frequently by the regularized Rolling Z-score (0.580), at a higher alarm rate than the machine-learning methods. Gradual increases were not reliably detected by any method; their plain event detection rates were within the range expected from background alarms alone. Detection performance and alarm burden have to be considered together.

Keywords: Anomaly Detection; Energy Monitoring; Industrial Machine Time Series; Synthetic Anomalies; Unsupervised Learning

1 Scientific Background and Motivation

Electrical energy consumption is an important economic and environmental factor in industrial production. In addition to total consumption, demand-related grid charges may depend on the maximum power demand reached within a billing period [1], so that unusually high load peaks are relevant for industrial energy management. With a growing number of monitored machines and changing machine behavior, manual evaluation and fixed thresholds become impractical, and automated methods are needed to identify isolated peaks, sudden shifts, or persistent changes [2], while a high power value is not anomalous if it is part of normal operation. As published studies differ in data and evaluation procedures and labels are rarely available, this work compares interpretable statistical approaches with unsupervised machine-learning methods on real industrial machine time series, using synthetic anomalies at known positions and explicitly accounting for the alarm burden and for detections expected from background alarms alone.

Anomaly detection aims to identify observations or patterns that deviate from an expected notion of normal behavior [3]; in time series, anomalies are commonly distinguished as point, contextual, or collective anomalies [4]. Here, detected anomalies are interpreted as unusual consumption values or patterns that require further examination rather than as confirmed faults. Simple statistical methods provide transparent baselines. A classical Z-score measures the deviation of an observation from a reference mean in units of standard deviation and is defined as

$$z_t = \frac{x_t - \mu}{\sigma},$$

where x_t denotes the observation at time t, while μ and σ represent the reference mean and standard deviation, respectively. In a rolling formulation, the reference statistics are calculated within a local time window rather than over the entire reference series, allowing the method to adapt to changes in machine behavior over time. Because

the mean and standard deviation can be strongly influenced by extreme observations, a more robust alternative is the Modified Z-score, which replaces them with the median and the Median Absolute Deviation (MAD):

$$MAD = \text{median}(|\, x_i - \tilde{x}\, |)\,, \qquad M_i = 0.6745\frac{x_i - \tilde{x}}{MAD},$$

where $\tilde{x}$ denotes the sample median. The same robust principle can also be applied within a moving time window, resulting in a Rolling Modified Z-score that combines robustness with local adaptation.

Unsupervised methods are relevant when anomaly labels are unavailable. Isolation Forest isolates observations by recursive random partitioning of the feature space; observations isolated with few partitions receive high anomaly scores [5]. Local Outlier Factor (LOF) compares the local density of an observation with that of its neighbors and can identify locally unusual observations that are not extreme globally [6].

Local statistical models have been applied to industrial time series [2] and LOF to energy data of a processing factory [7]; a review for buildings names missing labels and heterogeneous evaluation procedures as central obstacles [8]. Large comparative evaluations indicate that no single detector dominates across anomaly types and that simple methods remain competitive [9]. The evaluation protocols themselves have been criticized: common benchmarks contain trivial or mislabeled anomalies [10], and point-adjusted scoring, in which one alarm inside a segment counts the whole segment as detected, makes even random scores appear competitive [11]. The present study therefore reports detection per anomaly type together with the alarm burden, comparing event-level detection with the rate expected from background alarms, and uses a paired measure that isolates the effect of the injected change.

2 Materials and Methods - Data Acquisition and Edge Aggregation

The data were collected in two manufacturing companies in which 73 machines are monitored (milling, tempering, hardening, grinding, straightening, saw-blade setting, welding, sawing, compressors). Three-phase energy meters (Shelly Pro 3EM) in the sub-distribution boards measure the phase voltages directly and the currents via clip-on current transformers; active power and further quantities are computed in the meter and sent via MQTT to a local server in each company. Raw values at one-second resolution (600 per ten-minute interval) are reduced by an edge pipeline to 25 statistical features per window (mean, minimum, maximum, standard deviation, RMS, crest factor, skewness, kurtosis, entropy, rates of change, among others). Fig. 1 summarizes the workflow.

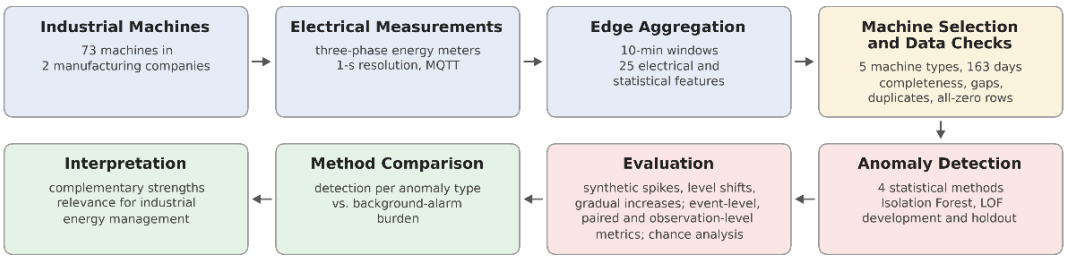


Fig. 1. Overview of the methodological workflow

3 Materials and Methods - Machine Selection, Analysis Period, and Data Quality

Since several machines of the same type are operated, five machine types were selected before the methods were applied and one machine per type was used: a carbide grinding machine (HM_Schleifmaschine_2_3), a milling machine (Fraesen_3), a straightening machine (Richten_1), a tempering furnace (Anlassen_2), and a saw-blade setting machine (Schraenken_2). The selection covers different profiles: Richten_1 is stable (median about 330 W, interquartile range about 323 W to 377 W); Fraesen_3 has widely spread operating levels (central 50% between about 772 W and 2834 W); Anlassen_2 combines a median of about 298 W with regular window means of up to about 93.3 kW. Data from 1 February to 13 July 2026 were used, i.e. about 163 days and 23,000 observations per machine (116,150 observations, about 815 machine-days). The series contained no missing values, duplicates, or negative electrical

values; 99.9% of the spacings lay between nine and eleven minutes, and no gap exceeded 20 minutes. Rows with all features equal to zero are frequent in some of the 73 machines but occurred only 33 times (at most 0.068% per series) in the selected series; they were kept as input but excluded from scoring and from the placement of synthetic anomalies. Low-power idle periods were treated as regular observations by all methods. The mean active power per window (Total Actual Power (W) Mean) was the only input signal, because it is directly related to the investigated patterns; the series are thus univariate, and the value of the remaining features was not assessed.

4 Materials and Methods - Applied Anomaly-Detection Methods and Method Development

Six approaches were compared (cf. Tab. 1): Global Z-score, Rolling Z-score, Global Modified Z-score, and Rolling Modified Z-score as interpretable baselines as well as Isolation Forest and LOF. Because the focus is on unusually high consumption, the statistical rules were applied in the upward direction only. A separate development stage covered, for the rolling methods, the reference window, the threshold, and two regularization measures: a minimum number of reference observations before a score is computed, and a local scale floor, by which the rolling standard deviation or MAD is replaced by 2% of the current rolling mean or median whenever it falls below that value. Without the floor, nearly constant reference windows during idle operation turn negligible changes into alarms.

Tab. 1. Final configurations of the anomaly-detection methods

Method	**Features**	**Final configuration**
Global Z-score	Power	$z > 3$; whole-series mean and SD (non-causal)
Rolling Z-score	Power	Past-only 1.5 h reference (9 obs.); $z > 3.70$; 2% local scale floor; min. 7 reference obs.
Global Modified Z-score	Power	$M > 3.5$; whole-series median and MAD (non-causal)
Rolling Modified Z-score	Power	Past-only 1 h reference (6 obs.); $M > 6$; 2% local MAD floor; min. 5 reference obs.
Isolation Forest	Power, ΔPower, z_3h	500 trees; max_samples = 1024; fitted per machine; top 1% score budget; upward-context filter
LOF	Power, z_3h	k = 50; no feature scaling; fitted per machine; top 1% score budget; upward-context filter

Note. ΔPower denotes the change from the preceding observation; z_3h is the deviation of the current power from a causal three-hour reference (rolling mean and standard deviation of the preceding 18 observations).

After investigating feature representations, model parameters, feature scaling, and score thresholds, both methods flag the observations with the top 1% of anomaly scores of a machine series, and the upward-context filter retains an alarm only when the power of the flagged observation exceeds the median of the preceding three hours. Both models were fitted separately per machine on the complete series and scored on the same series (transductive, without a separate training period); in the paired evaluation, they were refitted on the injected copy, so the injected observations were part of the fitted data. Isolation Forest used a fixed random seed. LOF uses unscaled features because scaled variants did not consistently improve the development trade-off; since power (W) and z_3h (dimensionless) differ by orders of magnitude, the distances are dominated by power, and this LOF variant effectively acts as a density-based detector on the power level. One configuration per method was fixed on the development benchmark before the holdout evaluation.

5 Materials and Methods - Synthetic Anomaly Generation

Three anomaly types were inserted into copies of the power series. A short load spike increases a single window by one or two times the 99th percentile of the machine's positive power differences (moderate and strong variant), capped at 1.5 times the historical non-zero maximum. A sudden level shift raises the level for 36 windows (about six hours) by 0.25 or 0.50 IQR of the machine. A gradual increase is a linear ramp over 144 windows (about 24 hours) reaching the same levels. Positions required sufficient preceding context and no overlap with all-zero periods, gaps,

or other events; magnitudes were defined per machine, and the temporal features were recomputed from the modified series. In the holdout benchmark, 400 spikes, 300 level shifts, and 60 gradual increases were injected across the eligible machine series using five seeds, with equal numbers of moderate and strong events for each anomaly type. The magnitudes represent noticeable but not gross deviations; a shift of 0.25 to 0.50 IQR corresponds to about 14 W to 27 W for Richten_1 and 0.5 kW to 1.0 kW for Fraesen_3.

6 Materials and Methods - Evaluation Metrics, Chance Analysis, and Holdout Protocol

At the observation level, precision is the share of flagged observations within the evaluation range that coincide with injected points, recall the share of injected points that were flagged, and F1 their harmonic mean. Alarms outside the injected intervals may correspond to unknown real unusual behavior and are referred to as background alarms rather than confirmed false positives; precision and F1 are therefore lower bounds. Their burden is expressed as alarm points per machine-day on the original series. At the event level, the event detection rate (EDR) is the share of injected events with at least one alarm inside the event. Paired incremental event detection compares the alarms on the original and the injected series at the same positions: an event counts as detected if at least one alarm occurs at a position inside the event that was alarm-free in the original (events already fully flagged in the original are excluded from the denominator). This primary measure isolates the effect of the injected change; for the machine-learning methods, refitting on the injected series can shift alarms, so a small part of their incremental detections may result from re-ranking. The EDR of a long event is inflated by background alarms: a randomly placed interval of L observations in a series with alarm fraction p contains at least one alarm with probability $1 - (1 - p)^L$ if alarms were independent. This chance hit rate is reported next to the EDR; as background alarms are clustered, it is an upper reference, and the paired measure is the empirical control. The holdout used new seeds and positions with the same shapes, magnitudes, durations, and placement rules and served exclusively for evaluating the fixed configurations. A three-hour post-event buffer was excluded from background-alarm scoring.

7 Results - Final Holdout Performance and Chance Analysis

Tab. 2 summarizes the holdout results. For short load spikes, Isolation Forest (0.900), the regularized Rolling Modified Z-score (0.875), the regularized Rolling Z-score (0.841), and LOF (0.810) detected the large majority of events; the differences are small and, without a measure of uncertainty for the pooled rates, are not interpreted as a ranking. For the six-hour level shift, the EDRs of the rolling methods, Isolation Forest, and LOF exceed their chance references, and the paired rates (0.580, 0.470, 0.377, 0.343) confirm that most detections were caused by the injected change. For the 24-hour gradual increase, the chance references lie between 0.478 and about 1.0, and the EDRs (0.08 to 0.55) are at or below them for every method, which also shows that background alarms are temporally clustered. The paired analysis shows that only few events produced a new alarm: 0.01 for the Rolling Z-score and 0.04 for the Rolling Modified Z-score despite EDRs of 0.55 and 0.52, i.e. almost all alarms inside the ramps would have been raised anyway. The highest paired rates, 0.310 for LOF and 0.260 for the Global Modified Z-score, remain low; for the latter, the incremental point detection of 0.092 indicates that the ramp pushes windows above the global threshold towards its end, and for LOF the re-ranking caveat applies. The result for gradual increases is therefore described as weak incremental sensitivity rather than detection. Across the five machines, paired spike detection ranged from 0.76 to 0.91 (Rolling Z), 0.72 to 0.97 (Rolling Modified Z), 0.81 to 0.95 (Isolation Forest), and 0.58 to 0.96 (LOF), with overlapping ranges, whereas the global methods varied from 0.00 to 0.66 (Global Z) and 0.01 to 1.00 (Global Modified Z). Level-shift detection varied more widely (e.g. 0.40 to 0.73 for Rolling Z, 0.00 to 0.82 for Isolation Forest), and the alarm burden ranged from 1.06 to 1.55 (Rolling Z) and 0.46 to 1.36 (Isolation Forest) to 0.00 to 53.8 alarms per machine-day for the Global Modified Z-score, whose pooled rate is dominated by single machines.

Tab. 2. Final holdout results: paired incremental event detection, EDR with chance reference, and alarm burden

Method	Spike	Level shift (36 windows)		Gradual increase (144 windows)		Alarms / Machine-Day
	Paired	Paired	EDR / chance	Paired	EDR / chance	
Global Z	0.205	0.040	0.057 / 0.242	0.020	0.080 / 0.670	1.106
Reg. Rolling Z	0.841	**0.580**	0.683 / 0.281	0.010	0.550 / 0.732	1.312
Global Modified Z	0.683	0.231	0.293 / 0.996	0.260	0.400 / 1.000	20.447
Reg. Rolling Modified Z	0.875	0.470	0.630 / 0.568	0.040	0.520 / 0.965	3.318
Isolation Forest	**0.900**	0.377	0.427 / 0.183	0.140	0.320 / 0.555	0.806
LOF	0.810	0.343	0.370 / 0.150	**0.310**	0.490 / 0.478	**0.649**

Note. Paired: paired incremental event detection rate. EDR / chance: plain event detection rate and expected hit rate of a randomly placed interval of the same length under independent alarms, $1 - (1 - p)^L$, with p the alarm fraction (alarms per machine-day divided by 144) and L = 36 or 144 windows. Alarms per machine-day refer to the original series. Best paired value per column and lowest alarm burden in bold.

8 Results - Observation-Level Precision, Recall, and F1

Tab. 3 reports the observation-level metrics. For spikes, recall equals the EDR because each event occupies one window. Precision is low for all methods (0.03 to 0.26) because every background alarm within the evaluation range counts against it, and it is not monotonically related to the alarm rate. For level shifts and gradual increases, recall is low for all methods except the Global Modified Z-score (0.252 and 0.196), whereas the EDRs are considerably higher: the rolling methods react at the onset of a sustained increase and then absorb the new level into their reference, and the machine-learning methods flag isolated windows within an event, so a single alarm inside an event documents neither its full duration nor its early detection. The higher recall of the Global Modified Z-score follows from its alarm fraction of about 14%. Because precision is bounded by the unlabeled background data, these metrics are not the decisive selection criterion.

Tab. 3. Observation-level precision (P), recall (R), and F1 on the final holdout benchmark

Method	Spike			Level shift			Gradual increase		
	P	R	F1	P	R	F1	P	R	F1
Global Z	0.256	0.210	0.195	0.017	0.008	0.009	0.006	0.007	0.005
Reg. Rolling Z	0.038	0.843	0.074	0.024	0.024	0.024	0.008	0.009	0.008
Global Modified Z	0.259	0.685	0.217	0.043	0.252	0.016	0.026	0.196	0.012
Reg. Rolling Modified Z	0.029	0.875	0.055	0.022	0.055	0.028	0.006	0.022	0.009
Isolation Forest	0.084	0.900	0.154	0.034	0.016	0.022	0.011	0.007	0.009
LOF	0.066	0.810	0.121	0.046	0.015	0.021	0.076	0.029	0.037

Note. Computed against the injection labels; precision is reduced by all background alarms within the evaluation range. Values are means over the evaluation units of the holdout benchmark; F1 was computed per unit and then averaged and therefore does not equal the harmonic mean of the tabulated P and R.

The initial Isolation Forest configuration without the causal temporal features and the upward-context condition reached spike and level-shift detection of 0.268 and 0.280 at 1.423 alarms per machine-day; with ΔPower, z_3h, and the upward-context condition it reached 0.894 and 0.381 at 0.797. This comparison also serves as an ablation: a large part of the advantage of Isolation Forest comes from the causal features (z_3h is itself a rolling statistical score) and the filter rather than from the model alone; as both were introduced together, their contributions cannot be separated. The original Rolling Modified Z-score produced 8.03 alarms per machine-day; the MAD floor and a shorter window reduced this to 2.83 at a spike detection of 0.879 compared with 0.894. The regularized Rolling Z-score reduced the alarm burden by about 54% while slightly improving spike detection. For LOF, spike detection increased from 0.390 to 0.815 and the alarm burden decreased from 1.435 to 0.649.

9 Discussion and Conclusions

The suitability of a method depends strongly on the type of consumption pattern. For spikes, Isolation Forest, the two regularized rolling methods, and LOF detected 81% to 90% of the events; the differences do not justify a ranking, but the machine-learning methods reached this level at the lowest alarm rates, which is their main practical advantage. For level shifts, the regularized Rolling Z-score detected the largest share, at about 1.6 times the alarm rate of Isolation Forest. The gradual increase was not reliably detected by any configuration, because a rolling window of six to nine observations absorbs a 24-hour ramp almost completely: the change within one window is only 4% to 6% of the final ramp magnitude (0.01 to 0.03 IQR), far below any threshold. Methods that accumulate evidence over time (CUSUM, EWMA [12, 13]) or seasonal references (same shift of previous weeks) are the natural candidates for this anomaly type and were not part of the comparison. Since the alarm rates range from 0.65 to 20.4 per machine-day, detection and alarm burden must be read together, and a comparison at matched budgets or as detection-versus-burden curves is the most important methodological extension. Without the paired measure, the EDRs for gradual increases (up to 0.55) would have suggested a sensitivity the methods do not have, an instance of the problem of point-adjusted scoring [11]. The high alarm rate of the Global Modified Z-score (14% of all windows) is most plausibly explained by heavy-tailed power distributions (Anlassen_2: median about 298 W, maximum about 93 kW), for which the global MAD is small compared with the power during regular load phases, so that these phases are flagged as a whole. A fixed global reference is therefore unsuitable for machines alternating between idle and load, whereas the rolling methods are protected by the scale floor.

Machine-level detection of unusual peaks and sustained increases can provide a basis for analyzing avoidable consumption, but the relation to demand-related grid charges is indirect. For customers with interval metering, the billing-relevant peak is the highest average power over a measurement interval, typically 15 minutes, at the grid connection point of the site [1], whereas the detections here refer to ten-minute means of single machines. Whether a machine-level alarm coincides with a billing peak depends on the simultaneous load of all other consumers, and a ten-minute spike is diluted in the 15-minute average. The results show that unusual machine-level patterns can be identified with a manageable alarm burden; they do not quantify a peak-shaving potential, which would require site-level aggregation aligned with the billing intervals.

The quantitative evaluation is based on synthetic anomalies. Development and holdout share the same anomaly generator, and the thresholds of the rolling methods ($z > 3.70$, $M > 6$) were tuned on the development benchmark; the holdout therefore shows generalization to new positions and seeds of the same shapes, not transfer to real anomalies. A validation against real events confirmed by experts is the most important next step. Results are reported without per-seed or per-magnitude breakdowns or confidence intervals, which would allow firmer statements about the small differences between the best methods. The comparison was not performed at matched alarm budgets, the five machines do not represent the population of 73, and the global methods are not causal. Production states, orders, or maintenance information was not available, which limits the evaluation of contextual anomalies, and the ten-minute aggregation hides changes within one interval. Future work should include such operational information, the remaining electrical features, change-detection and seasonal references for gradual increases, and a matched-budget comparison. Despite these limitations, simple statistical methods and selected machine-learning approaches can support the identification of unusual industrial consumption patterns with a manageable alarm burden, provided that the anomaly type of interest and the acceptable alarm rate are specified beforehand.

References

[1] German Electricity Network Charges Ordinance (Stromnetzentgeltverordnung – StromNEV), § 17(2): Gesetze im Internet, URL: https://www.gesetze-im-internet.de/stromnev/__17.html, Accessed: 19.09.2026

[2] Theumer, P., Zeiser, R., Trauner, L., Reinhart, G. (2021): Anomaly Detection on Industrial Time Series for Retaining Energy Efficiency, Procedia CIRP 99, p. 33-38, doi: 10.1016/j.procir.2021.03.006

[3] Chandola, V., Banerjee, A., Kumar, V. (2009): Anomaly detection: A survey, ACM Computing Surveys 41(3), p. 1-58, doi: 10.1145/1541880.1541882

[4] Blázquez-García, A., Conde, A., Mori, U., Lozano, J. A. (2021): A Review on Outlier/Anomaly Detection in Time Series Data, ACM Computing Surveys 54(3), p. 1-33, doi: 10.1145/3444690

[5] Liu, F. T., Ting, K. M., Zhou, Z.-H. (2008): Isolation Forest, 2008 Eighth IEEE International Conference on Data Mining, p. 413-422, doi: 10.1109/ICDM.2008.17

[6] Breunig, M. M., Kriegel, H.-P., Ng, R. T., Sander, J. (2000): LOF: Identifying Density-Based Local Outliers, Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, p. 93-104, doi: 10.1145/342009.335388

[7] Dang, H.-A., Dao, V.-D., Dang, C.-D., Nguyen, N. H., Hoang, A. (2023): Identifying Abnormal Energy Consumption Patterns in Industrial Settings: Application of Local Outlier Factor Algorithm for a Processing Factory in Vietnam, 2023 Asia Meeting on Environment and Electrical Engineering (EEE-AM), p. 1-5, doi: 10.1109/EEE-AM58328.2023.10395078

[8] Himeur, Y., Ghanem, K., Alsalemi, A., Bensaali, F., Amira, A. (2021): Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives, Applied Energy 287, 116601, doi: 10.1016/j.apenergy.2021.116601

[9] Schmidl, S., Wenig, P., Papenbrock, T. (2022): Anomaly Detection in Time Series: A Comprehensive Evaluation, Proceedings of the VLDB Endowment 15(9), p. 1779-1797, doi: 10.14778/3538598.3538602

[10] Wu, R., Keogh, E. J. (2023): Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress, IEEE Transactions on Knowledge and Data Engineering 35(3), p. 2421-2429, doi: 10.1109/TKDE.2021.3112126

[11] Kim, S., Choi, K., Choi, H.-S., Lee, B., Yoon, S. (2022): Towards a Rigorous Evaluation of Time-Series Anomaly Detection, Proceedings of the AAAI Conference on Artificial Intelligence 36(7), p. 7194-7201, doi: 10.1609/aaai.v36i7.20680

[12] Page, E. S. (1954): Continuous Inspection Schemes, Biometrika 41(1/2), p. 100-115, doi: 10.1093/biomet/41.1-2.100

[13] Roberts, S. W. (1959): Control Chart Tests Based on Geometric Moving Averages, Technometrics 1(3), p. 239-250, doi: 10.1080/00401706.1959.10489860

From LSTM Baselines to Edge-Deployed Mixture-of-Experts: Lessons from Fielded Neural Automatic Passenger Counting

Dastgheib Shirazi, Ramtin (ramtin.shirazi@interautomation.de),
Schickel, Yannick (yannick.schickel@interautomation.de)
Interautomation Deutschland GmbH, Hauptstr. 56 – 60, 13158 Berlin,
https://www.interautomation.de/en/

Abstract: Automatic passenger counting (APC) underpins fare clearing, capacity planning and real-time passenger information. Since the first prototype in 2020 our Neural APC pipeline has grown into a system fielded across 26 rolling-stock projects and three sensor suppliers, on door units with no GPU or NPU and, on the stereo-vision platform, no floating-point arithmetic at all. No persistent benchmark was ever maintained, so we treat the archives as the experiment. From more than 2,500 archived runs, each stored with its configuration, split and weights, we build one held-out set of 9,545 recordings absent from every split of the 76 runs compared, render it at each generation's own input resolution, and score all three generations against the acceptance criterion of VDV-Schrift 457. At a shared training budget and on the same training data, the 2021 convolutional front-end cuts per-video error by a fifth with 46 % fewer parameters, but it leaves counting bias where it was. The 2025 Mixture-of-Experts, whose experts build on that previous generation, brings every run inside the one-percent bias bound, also on its predecessor's training data, while evaluating only the part of itself its gate selects. Static ensembles, the arrangement we fielded before 2025, recover part of that bias but evaluate every one of their members. Fixed-point conversion, required by the stereo-vision sensor, costs an undercount governed mainly by the allowed per-frame error.

Keywords: automatic passenger counting, mixture of experts, sparse gating, edge inference, fixed-point quantization, retrospective evaluation, depth sensors

1 Introduction

Accurate automatic passenger counting is a commercial necessity rather than a convenience: it feeds fare clearing between operators, capacity planning and real-time occupancy displays, the application fields the industry specification sets out [1]. Counting from depth data is attractive because a depth sensor records no recognizable image of a passenger, which makes the method privacy-friendly by construction [2], and because a learned model can use the shape and motion of a whole scene rather than the interruption of a beam. The price is a model that must run in real time on the sensor itself, on units with no GPU or NPU and tight bounds on memory, clock and power. Every architectural decision reported here was made under that constraint, and much of what follows is how the constraint relaxed as the hardware matured.

Our pipeline originates in "vericon", a joint research project of TU Berlin and Interautomation Deutschland GmbH funded by the IBB Pro FIT program, which ran from May 2019 to April 2021. Since then, architecture, data and deployment targets have all changed substantially, while the pipeline around them stayed recognizably the same. Architectures were compared on shared training sets within a project, and a change that proved itself was carried forward, but no benchmark was kept that every later change was re-tested against, and label revisions, resolutions and hold-out definitions all moved between projects. That is the configuration and data debt Sculley et al. [3] describe as the characteristic long-run cost of a production system, and it is why no frozen reference point exists. This paper constructs one from the archive.

Neural counting on privacy-friendly depth data was established by Seidel et al. [2], who framed the task as regressing cumulative boarding and alighting counts from a door-mounted sensor sequence, and by Jahn and Siebert [4], who studied the engineering around it. Our pipeline descends directly from that work: the first NAPC model was a bachelor thesis inside vericon, the convolutional generation a master thesis submitted at TU Berlin in February 2021 by one of the present authors.

What distinguishes this application domain from most machine-learning practice is that the acceptance criterion is fixed externally. VDV-Schrift 457 [1], the German industry specification for automatic passenger counting, does not accept a system on mean absolute error but on *bias*. Section 7.1 requires that the summed difference between automatic and manual counts, over all stops of all test runs and per counting-vehicle category, stay within ±1 % of the manual total, and that unbiasedness be demonstrated with an equivalence test whose permissible range is ±1.5 %. The procedure is the two one-sided tests of Schuirmann [5], which reverse the burden of proof onto demonstrating agreement. The sample must cover at least 6,000 door-stop events. We express the same quantities per direction as a relative bias $\bar{D}$, the mean signed count error over the mean manual count, and a relative spread ν. Both enter the equivalence margin

$$m = \Delta - z(1-\alpha/2) \cdot v / \sqrt{n} - |\bar{D} - \bar{D}_0|$$

where $\bar{D}_0$ is the bias tested against, normally zero. The margin is non-negative exactly when the equivalence test passes, and it combines bias, spread and sample size in one number. We test against $\Delta = 0.01$ with $\alpha = 0.05$, stricter than the 0.015 the specification permits, and we release a model only when the two-sided t-test passes as well. The equivalence test binds tenders published from 30 June 2019 onwards, and the specification does not recommend applying it retroactively [1], so campaigns on our existing fleet still use the t-test. Two properties of the criterion shape what follows. A one-percent bound is tight enough that fixed-point rounding matters, and a signed bias counted per direction can look small through cancellation.

2 Methods

2.1 The pipeline and its three architecture generations

The NAPC pipeline is a configuration-driven chain of five components: a video handler that reads the recordings of each sensor generation in whichever format they were stored, a data stage building memory-mapped tensors and labels, a training stage whose network is specified declaratively layer by layer in a configuration file, a converter that searches fixed-point scales and emits C++, and an evaluation stage scoring predictions statistically. Two properties carry this study: the architecture is data rather than code, so a 2020 and a 2026 network are two files in the same format, and every run archives its configuration, split and weights together.

Sensor frames arrive at 160 × 120 pixels and are reduced by the data stage. The first NAPC model (2020, bachelor thesis) worked on 25 × 20 frames, flattening each one through a dense projection and five stacked LSTM layers into cumulative counts, which discarded almost all spatial structure. The 2021 master thesis revisited the bounding-box labels and inserted three convolution and pooling stages ahead of a shortened LSTM stack. That let the network learn object shape and position rather than one generic "person" signature, which is also what made counting bicycles carried through the doors possible. Later generations moved to larger inputs, 40 × 30 and 40 × 35.

2.2 Why static ensembles gave way to sparse routing

Before 2025, variability between projects was handled with static ensembles of independently trained CNN-LSTM models, selected per project. Rather than averaging their outputs, the four members were sorted per direction and the second highest count was taken. That worked, but at a price. Several models had to be run and aggregated on already constrained hardware, and every deployment needed its own evaluation and selection procedure, an overhead that scaled poorly as projects multiplied. Project-specific training carried a subtler limitation. Many operationally relevant events, such as unusual flows, bicycles, wheelchairs, strollers, large luggage or the installation quirks of one vehicle, are rare in any single project and stay underrepresented in its training data, while the recordings from all deployments together form a far richer corpus that a per-project pipeline cannot draw on.

Our experiments then identified why static ensembles cannot exploit such events even when one member has seen them. Jacobs et al. [6] drew the relevant distinction when they introduced mixtures of experts: combining models that all answer is not the same as selecting the one whose competence the input falls into. Because every member contributes to every prediction and the outputs are pooled, pooling systematically suppresses rare events. A model that has learned to recognize bicycle passages or unusually high counts is outvoted at inference by members that never saw comparable data and fall back on the common prediction. The event is detected by one member and lost in aggregation. This is not a training deficiency but a structural property, since a prediction is diluted in proportion to how many models disagree with it, and rare events are by definition what most models disagree on.

Sparse routing [7] removes that failure mode, because it does not require agreement. The gating network activates only the experts it judges relevant, and unselected experts have no influence at all. Routing happens per sequence, and only the selected experts are evaluated, the sparse regime Fedus et al. [8] showed to be sufficient and the cheapest to evaluate. A rare event therefore does not need to convince a majority. It only needs to reach the expert that specializes in it. In our evaluations the MoE detected bicycle passages and unusually high counts that the static ensemble missed, without losing accuracy on common cases. Because the experts train jointly on data pooled from all deployments, specialization forms around recurring operational patterns rather than project identity, and a generalist capacity emerges for the ordinary passenger motion that nearly all projects share. That is what let predictions on newly onboarded projects hold up before any fine-tuning.

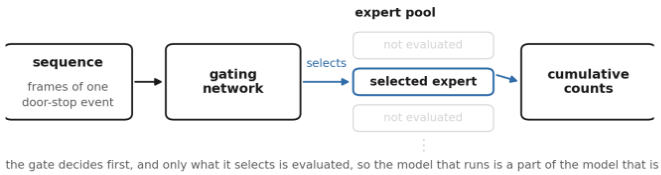


Figure 1 The Mixture-of-Experts as deployed. The gating network scores the expert pool for each sequence and only the experts it selects are evaluated, so the model that runs is a part of the model that is stored. The experts build on the convolutional architecture of the previous generation.

2.3 What the sensor generations allowed

The platforms differ in what they can execute, and that drove the numeric format far more than any modeling consideration. The HAPC-R derives depth from stereo vision and offers no floating-point arithmetic, so a model could only reach it as integer-arithmetic-only code in the sense of Jacob et al. [9]. The first time-of-flight generation, Dist500, could not execute a model binary on the sensor at all, so its quantized binary ran on an on-board computer with no accelerator either. Only the second, Dist80T, holds several floating-point models on the sensor itself, which removed quantization, made ensembles practical and let us collect labeled recordings at scale from 2022. Figure **2** reads that progression out of the compiled binaries. The models now in service are fine-tuned from a baseline trained on a large time-of-flight corpus.

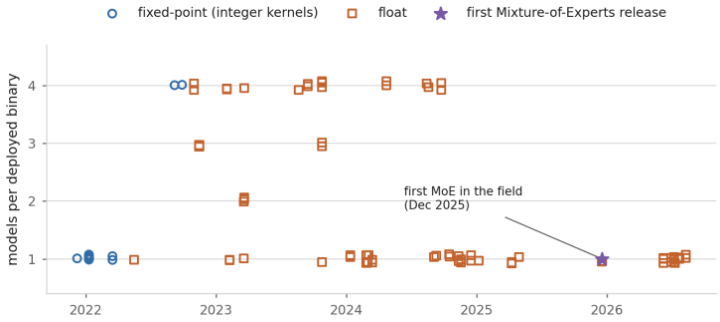


Figure 2 The 84 deployed artifacts over time: how many models each holds, and whether it was converted to fixed point. Binaries holding more than one model are the static ensembles of Section 2.2, and the first mixture is the marked release.

Across all generations, trained models are converted into hand-optimized C++ with custom kernels for the convolutional and recurrent operations, and since 2025 for gating and routing as well, targeted at each platform's instruction set and memory layout. Every converted model is validated end-to-end against its floating-point counterpart on a hold-out set before it may be deployed.

2.4 How model selection matured

Model selection followed the same trajectory as the hardware. In the first years the labeled corpus was small, so the team reserved 4,000 to 6,000 recordings as a test set, the scale of a real verification sample, and trained on the rest. With nothing to spare for a validation split, models ran a fixed 6,000 epochs and the final epoch was carried forward, then quantized for the platforms of Section 2.3. From 2022, with Dist80T supplying recordings across many projects, a validation split cost nothing. We validated during training and kept the best epoch by lowest validation loss, which combined well with ensembles and remains our default. Where those weights could not be used for a release, we evaluated a model's epochs in the inference component and selected on test-set performance instead.

For the Mixture-of-Experts we kept validation loss and experimented with selecting on the validation bias directly. That did not yield better models, and Section 3.3 shows why a signed, direction-averaged bias is a weak signal. Since the migration to a CI/CD pipeline, validation during training computes the acceptance criterion itself, the per-direction equivalence margin, rather than a proxy for it.

2.5 A retrospective evaluation protocol

The archive we evaluate holds more than 2,500 training runs carrying a configuration, 84 dated deployment artifacts across 26 projects, and archived prediction sets from 2022 onwards. Our protocol has five rules.

1. One metric implementation for everything, applied per direction, never averaged across directions, and applied to models from 2022 and 2026 alike rather than to what a run logged at the time.

2. One held-out set for every generation. A recording qualifies only if it appears in no training and no validation split of any of the 76 runs compared, carries the same label in every dataset version that mentions it, and still has its source file. Of more than 65,000 labeled recordings from the sensor family all three generations share, 9,545 qualify, which is above the 6,000 door-stop events the specification requires.
3. Every run is assigned to its generation by the layer graph stored in its checkpoint, so a row of the comparison contains the architecture it was actually trained as.
4. Every generation is read at a checkpoint the whole comparison shares. Training budgets in the archive are uneven, so comparing each run at its own last epoch would measure when a run was stopped. Each arm therefore fixes one epoch number that every run in it reached.
5. Every reproduction is checked before it is believed. A harness that aggregates only part of a sequence must undercount more the more events a video holds, so the ratio of predicted to true counts on videos with at least one event and with at least three must agree within 0.05.

3 Results

3.1 Architecture generations and their cost at the edge

The convolutional front-end of 2021 is the result most worth passing on: it improved counting accuracy while making the network *smaller*. It uses 87,400 parameters against the 161,152 of the recurrent stack it replaced, 46 % fewer. Three convolution stages remove more parameters than they add by replacing a flattened-frame dense projection, and recurrent depth becomes unnecessary once features carry spatial structure.

The 2025 mixture moves the other way on storage, since its experts build on the convolutional architecture of the previous generation and holding several of them makes the stored model a multiple of a single one. What runs on the sensor is not the stored model. Only the experts the gate selects are evaluated, so the arithmetic stays a small multiple of one previous-generation model rather than the whole pool. That is what makes the mixture admissible on a sensor with no accelerator, where the binding constraint is arithmetic per frame and flash is comparatively cheap. It is also less arithmetic than the static ensembles it replaced, which evaluate every member on every recording.

3.2 Counting accuracy on recordings no model has seen

All three generations exist on one sensor family at two input resolutions drawn from the same recording campaign, so one held-out set serves all of them: the 9,545 recordings of Section 2.5, rendered twice so every model is scored on the same recordings at the input size it was built for. The mean manual count per video is 1.71.

Generation	Input	Models	Epochs	D̄ boarding	D̄ alighting	MAE boarding	MAE alighting	ν boarding	Within ±1 %
Stacked LSTM, 5 layers	40 × 30	8	6,000	−0.0395	−0.0296	0.099	0.087	0.330	0/8
Stacked LSTM, 1 layer	40 × 30	26	6,000	−0.0200	−0.0183	0.085	0.080	0.282	0/26
CNN-LSTM	40 × 30	12	6,000	−0.0207	−0.0171	0.069	0.068	0.257	0/12
CNN-LSTM	40 × 35	6	2,000	−0.0141	−0.0107	0.068	0.065	0.249	2/6
Ensemble of three, middle	40 × 35	20	2,000	−0.0156	−0.0133	0.062	0.058	0.239	0/20
Ensemble of four, second highest	40 × 35	15	2,000	−0.0076	−0.0064	0.060	0.056	0.225	13/15
Mixture-of-Experts	40 × 35	6	2,000	−0.0041	−0.0007	0.058	0.056	0.196	6/6

Table 1 Every generation on the same 9,545 held-out recordings, each model read at the epoch its arm fixes. D̄ is the signed relative bias per direction, MAE the mean absolute count error per video and ν the relative spread the acceptance tests divide by, all medians over models. The last column counts models whose worse direction stays inside the ±1 % bound of Section 1. The two CNN-LSTM rows are the same architecture at the two resolutions, and bridge the comparison. The two ensemble rows are the aggregation rules of Section 2.2, covering every combination that can be formed from the 6 CNN-LSTM runs above them.

Two things improve, and they improve separately. Each run trained on the data of its time, so we read both on runs that share their training recordings. The 1-layer LSTM and CNN-LSTM runs at 40 × 30 share one split, and there the convolutional front-end cuts the per-video error by roughly a fifth, from a median MAE of 0.085 to 0.069 for boarding and from 0.080 to 0.068 for alighting, while leaving the relative bias essentially where it was. The mixture reaches a median D̄ of −0.0041 and −0.0007, against −0.0141 and −0.0107 for the CNN-LSTM trained to the same epoch, and 6 of 6 mixture runs stay inside the one-percent bound where 2 of 6 CNN-LSTM runs do. Four mixtures but only one CNN-LSTM trained on a larger pool. On the pool both generations share, the 2 mixtures still stay inside the bound and pass the equivalence test, against 2 and 1 of 5 CNN-LSTM runs, at the same median MAE of 0.070 for boarding. Routing moves the bias on matched data, and the lower error of the mixture row comes with the larger pool.

The static ensembles we fielded before 2025 sit between the two, and their two rules show where the benefit comes from. Taking the middle of three sorted members narrows the spread the acceptance tests divide by, from a median ν of 0.249 to 0.239 for boarding, but leaves the bias where it was. Taking the second highest of four narrows the spread further and also halves the bias, to −0.0076 and −0.0064, and 13 of those 15 ensembles stay inside the one-percent bound where none of the three-member ones do. These recombine 6 runs, so they are not 15 independent results. Leaving out one run at a time, 3 to 5 of the remaining 5 stay inside the bound. Both rules keep a count some model actually produced. Only the off-center one corrects a generation that undercounts, which is why it shipped. The mixture still ends up ahead of both, while evaluating only the experts its gate selects rather than every member.

The ordering survives a different cut. At the earlier matched budget of 4,000 epochs the convolutional front-end again lowers the per-video error without lowering bias. Read at their own last checkpoints instead, up to 5,300 epochs, no CNN-LSTM run passes the equivalence test while every mixture does. More training is not the difference, and it can cost: the best CNN-LSTM clears the boarding equivalence test at 2,000 epochs and fails it at its own last epoch.

3.3 Which acceptance test binds

Which of the two acceptance tests binds depends on the spread. On the held-out set of Section 3.2 the two tests bound the bias by

$$\Delta \;-\; z(1-\alpha/2)\cdot \nu / \sqrt{n} \qquad \text{and} \qquad z(1-\beta/2)\cdot \nu / \sqrt{n}$$

respectively, so with ν small relative to Δ the t-test is the tighter of the two. The mixture runs show it: 4 of 6 pass the equivalence test and 2 pass the t-test. Neither test is redundant, and neither ordering holds in general. The ensembles make the point from the other side. Thirteen of the 15 four-member ensembles keep their bias inside ±1 %, but only 3 pass the equivalence test, all lacking the same run, because even a rule built to narrow the spread leaves it near the level of its members (ν = 0.225 against 0.196 for the mixtures) and the confidence interval eats the room the bound leaves. A bias inside the bound is not the same as a demonstrated one. The margin shows how little a bare pass can mean: the best CNN-LSTM of Table **1** clears the equivalence test on boarding by +0.00002 against +0.00418 for the best mixture. Both pass, but only one has room for the next vehicle, sensor batch or labeling revision.

3.4 What fixed-point conversion cost

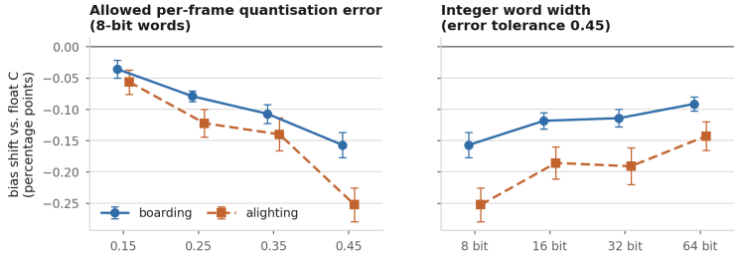


Figure 3 Bias shift induced by fixed-point conversion, relative to each model's own floating-point C reference, on 6,342 held-out videos. Markers are means over repeated quantization runs of five models (5 to 24 per configuration) and bars the standard error of those means. Both panels share the vertical scale.

Fixed point was a hardware requirement rather than a design preference, but until now we had never priced it. A 2022 sweep across scale-search settings, re-analyzed with the current test suite, separates the two knobs. The floating-point C implementation first reproduces TensorFlow to within 0.015 percentage points of relative bias, which validates the hand-written kernels. Both settings then matter, unequally. Loosening the allowed per-frame error from 0.15 to 0.45 at 8-bit words multiplies the induced undercount roughly fourfold, from −0.036 / −0.057 percentage points to −0.157 / −0.253. Widening the word from 8 to 64 bits at that tolerance recovers part of it, to −0.092 / −0.143, about a factor of 1.8. The effect is monotone in the word width and well outside the standard error of the means.

The tolerance is thus the dominant lever and the word width a secondary one, but in practice both ends were constrained: the widest words were too expensive to run on the target, and the tightest tolerances often found no valid scale. The configuration that shipped sat at the costly end of both axes because those were the only settings that produced a deployable model, and the residual undercount it carries is the price the hardware charged for having no floating-point unit.

4 Discussion

A selection metric should be the acceptance criterion, not a proxy. Our selection process improved in step with the data available: a fixed epoch count when nothing could be held back, best-epoch-by-validation-loss once Dist80T

supplied recordings at scale, and now the per-direction equivalence margin during training. Section 3.3 shows why the intermediate step of selecting on validation bias alone did not help, and why the margin is the right target. Table **1** shows why neither alternative suffices on its own: the convolutional generation improved loss and left bias untouched, while a signed bias can be driven toward zero by cancellation. Any program whose product is accepted on a statistical criterion should compute that criterion during training as early as it can afford to.

A long-lived archive needs its own engineering. Storing each run's configuration, split and weights together was the most valuable decision made here. Without it none of the results above would exist, and it is what let us evaluate models six years apart under one metric. Every claim in this paper is derived from the stored artifact rather than from what was recorded about it. The one thing we would add is a reference prediction archived with each run.

Threats to validity. The comparison controls the held-out recordings, the metric implementation and the training epoch. Training data is matched only where Section 3.2 says so, and label revisions not at all. A shared epoch count matches training passes, not compute. The mixture generation rests on 6 runs, 2 of them on matched data, and everything here comes from one vehicle family and one sensor type, the stereo-vision HAPC-R. The models now fielded are fine-tuned from a time-of-flight baseline this comparison does not cover. The ensemble rows of Table **1** reconstruct the aggregation rules we deployed, but their members are the runs of this archive rather than the per-project selections that were actually shipped. The rare-event and cross-project behavior of Section 2.2 rests on operational evaluations rather than a controlled experiment, and quantifying it on a common held-out set is the obvious next step.

5 Conclusion and outlook

Over six years the NAPC pipeline went from a stacked-LSTM prototype to a sparse Mixture-of-Experts on the door sensors of 26 rolling-stock projects across three suppliers, on hardware that began with no floating-point arithmetic at all. Treating more than 2,500 archived runs, 84 deployment artifacts and 9,545 never-seen recordings as the experiment let us quantify that trajectory without retraining anything, and the trajectory is not what a single accuracy figure would suggest. On matched data, convolution bought per-video error, routing bought bias, and only bias is what the specification grades. The ensembles in between bought part of that bias with an order statistic rather than with averaging, and by running every member rather than a selected one. Since December 2025, 17 of the 19 artifacts deployed were mixtures. Ongoing work reduces routing overhead on the most constrained platform, extends generalization to new sensor form factors, and completes the move to CI/CD and a cataloged data platform.

References

[1] S. Köhler, S. Bobinger, R. Branick, B.-M. Cerfontaine, und S. Krogull, *VDV-Schrift 457*, V2.2. Köln: VDV, 2023.

[2] R. Seidel, N. Jahn, S. Seo, T. Goerttler, und K. Obermayer, „NAPC: A Neural Algorithm for Automated Passenger Counting in Public Transport on a Privacy-Friendly Dataset", *IEEE Open Journal of Intelligent Transportation Systems*, Bd. 3, S. 33–44, 2022, doi: 10.1109/OJITS.2021.3139393.

[3] D. Sculley *u. a.*, „Hidden technical debt in Machine learning systems", in *Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 2*, in NIPS'15, vol. 2. Cambridge, MA, USA: MIT Press, Dez. 2015, S. 2503–2511.

[4] N. Jahn und M. Siebert, „Engineering the Neural Automatic Passenger Counter", *Engineering Applications of Artificial Intelligence*, Bd. 114, S. 105148, Sep. 2022, doi: 10.1016/j.engappai.2022.105148.

[5] D. J. Schuirmann, „A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability", *J Pharmacokinet Biopharm*, Bd. 15, Nr. 6, S. 657–680, Dez. 1987, doi: 10.1007/BF01068419.

[6] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, und G. E. Hinton, „Adaptive Mixtures of Local Experts", *Neural Computation*, Bd. 3, Nr. 1, S. 79–87, März 1991, doi: 10.1162/neco.1991.3.1.79.

[7] N. Shazeer *u. a.*, „Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer", 23. Januar 2017, *arXiv*: arXiv:1701.06538. doi: 10.48550/arXiv.1701.06538.

[8] W. Fedus, B. Zoph, und N. Shazeer, „Switch transformers: scaling to trillion parameter models with simple and efficient sparsity", *J. Mach. Learn. Res.*, Bd. 23, Nr. 1, S. 120:5232-120:5270, Jan. 2022.

[9] B. Jacob *u. a.*, „Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference", in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Juni 2018, S. 2704–2713. doi: 10.1109/CVPR.2018.00286.

RAISSE-Methodology for Interpretable Transfer Learning in Restricted Environments with Illustrative Examples

Thomas Pfaff[1], Marika Kaden[1], Thomas Villmann[1,2]
[1]Sächsisches Institut für Computational Intelligence und Machine Learning (SICIM)
Hochschule Mittweida, Technikumplatz 17, 09648 Mittweida
[2]Technische Universität Bergakademie Freiberg, Akademiestraße 6, 09599 Freiberg
thomas.pfaff@hs-mittweida.de, https://sicim.hs-mittweida.de

Abstract:

Deploying predictive machine learning models in restricted engineering environments frequently suffers from domain discrepancies caused by varying hardware setups, sensor tolerances, or environmental conditions. Many conventional domain adaptation approaches rely on complex representation-learning architectures or iterative target-dependent adaptation procedures, which can be computationally expensive, resource-intensive, and difficult to interpret. To address these challenges, the RAISSE (Reliable AI for Smart System Engineering) methodology shifts the domain adaptation paradigm from model adjustment to the refinement of interpretable representations. Rooted in prototype-based interpretable machine learning, RAISSE extends foundational concepts such as Generalized Relevance and Matrix Learning Vector Quantization (GRLVQ/GMLVQ) for industrial applications and specifically to domain adaptation scenarios. While feature relevance learning with GRLVQ successfully extracts the features that are important for a task and allows for a lot less sensors, there is still a risk of incorrectly identifying important setup-dependent artefacts. This work overcomes this limitation by coupling a domain-discriminative GMLVQ with Iterated Relevance Matrix Analysis (FA(IR)^{2}MA-GLVQ) to mitigate domain-specific distortion via null-space projection while aiming to preserve task-relevant structures. The resulting approach makes it possible, on the one hand, to apply a primary prediction model to unlabeled target data without requiring target task labels or retraining the model for the specific task. On the other hand, it also highlights when this is not possible, thereby fostering an understanding of the data and the issues involved and building trust in AI predictions. We validate the RAISSE methodology for transfer learning across two distinct real-world scenarios: (1) a gait phase recognition system adapting to a new environment with fewer sensors and without primary task labeled data, and (2) a subtle measurement setup shift in Near-Infrared (NIR) spectroscopy for biogas substrate analysis across multiple environment configurations. This work illustrates a transparent white-box paradigm for reliable, domain-robust transfer learning in industrial environments.

Keywords: interpretable ML; Transfer Learning; Classification; Bias Mitigation; Smart System Engineering

1 Scientific Background and Motivation

Transfer learning has emerged as a key paradigm in machine learning for deploying predictive models across varying acquisition, measurement, and operational conditions [1]. In real-world engineering model transferability is frequently hindered by domain-dependent systematic variations stemming from heterogeneous sensor configurations, hardware tolerances, environmental factors, or operational contexts. Conventional domain adaptation techniques typically attempt to overcome these discrepancies through feature alignment, domain-invariant representation learning, or fine-tuning network parameters [2]. An example is domain-adversarial learning, which explicitly aims to learn feature representations that are discriminative for the primary task while being invariant with respect to source and target domains [3]. While effective in many computer vision or general benchmark scenarios, these methods predominantly treat transferability as a model adaptation problem. They often act as black-box proce-

The authors would like to thank ICM e.V., Chemnitz, and Sächsischer Landeskontrollverband e.V., Niederwiesa, for experimental data and domain expertise. The projects were funded by 100376213 and 28DK112B20. T. P. is supported by BMFTR (PAL: 02L19C300, KIX: 13HAW076FQ) and M. K. supported by the European Regional Development (KI-Med, EFRE: 100734114)

dures, adapting the model iteratively to each target domain without providing explicit insights into the underlying causes of performance degradation. Such approaches face severe limitations in constrained environments, including limited instrumentation, missing ground truth labels in target domains, and computational constraints associated with complex model architectures. Moreover, applications involving safety- or decision-critical processes impose strong requirements on interpretability and reliable decision-making [4]. Constrained environments consequently require models that are inherently interpretable by design. To address these challenges, the RAISSE methodology (Reliable AI for Smart System Engineering) promotes the application of reliable, interpretable and mathematically well-defined AI systems to address engineering challenges like

- to obtain interpretable, robust analytic tools for system diagnostics/prediction/control, e.g. medical analytics, predictive maintenance, quality evaluation etc. with low-cost constraints for the AI-model
- integrate engineering and/or domain expert knowledge into the design of AI systems to achieve lower complexity and, hence, sparser hardware constraints to run an ML-powered diagnostic/control engineering system
- to identify smart hardware configurations in engineering, sensoric or measurement devices by AI-analytics
- transfer information from well-defined settings to new environments by means of reliable ML models

Previous work on relevance learning has demonstrated that interpretable metric adaptation, such as l_1-regularized Generalized Relevance Learning Vector Quantization (GRLVQ), can successfully enforce sparsity in high-dimensional spectral profiles, isolating a compact subset of highly discriminative bands (e.g., reducing hundreds of wavelengths to a few key spectral channels) while maintaining high classification accuracy [5], which provides substantial industrial value by simplifying sensor architectures and lowering hardware complexity. However, such relevance learning techniques implicitly assume domain homogeneity across training and deployment conditions [1]. When unmitigated setup-dependent or environmental shifts are present in the data representation, metric adaptation risks assigning high relevance to domain-specific artifacts rather than true physical characteristics [6].

In this work, we examine the foundational RAISSE principle for the systematic identification and mitigation of domain-dependent structures affecting transferability in restricted environments. By explicitly identifying and quantifying these structural biases, a discriminative null-space projection is applied to suppress domain-discriminative directions while aiming to preserve task-relevant information. By leveraging interpretable prototype-based machine learning, model transferability across restricted environments can be improved through representation cleaning, avoiding complex, domain-specific model fine-tuning in the investigated transfer setting. Ultimately, RAISSE enables AI models in constrained industrial settings to become not only robust against hardware biases and setup-induced drifts, but also highly resource-efficient by laying the groundwork for targeted sensor and feature reduction.

2 Methods

We suppose data from several measurement domains (sources) as n-dimensional (feature) vectors. These data required to be classified or to be evaluated by a regression model, which constitutes the primary classification/regression learning task. For classification/regression training a subset of training data only from one source is available, whereas for the data from all other sources only the measurements are available. Thus, transfer learning requires the detection of source discriminating information in the data and, if detected, to mitigate this information from the data as best as possible. Realizing the RAISSE principle for transfer learning we apply the Generalized Matrix Learning Vector Quantization (GMLVQ) as an interpretable prototype-based classifier model for the mitigation procedure followed by classifier training for the mitigated data.

GMLVQ builds upon the foundational works of Kohonen and Sato & Yamada [7, 8] for non-linear classification learning combined with metric learning concepts [9, 10]. The latter one considerably expanded the methodological capabilities and enabled successful applications providing information about feature relevance through its tasks adjusted model structure and is known as GRLVQ for scalar feature weighting and its matrix extension – GMLVQ [11, 12]. This inherent interpretability enables the examination of domain-related influences and classification decisions in a transparent manner, in contrast to conventional black-box classifier models like (deep) neural networks. Beyond transparency, prototype-based vector quantizers provide strong margins and verifiable boundary stability, resulting

in increased model robustness against signal noise and distribution shifts [13]. A key ingredient of relevance learning is that it is realized via the adaptation of a data embedding matrix (projection matrix) $\Omega \in \mathbb{R}^{m \times n}$, where n is the number of input features and m the dimension of the embedding space, such that the resulting relevance matrix $\Lambda = \Omega^T \Omega$ provides information about task-relevant feature structures and relations [11].

In our framework, we frame the transfer learning challenge as a representation problem, leveraging principles from fair machine learning, specifically the FA(IR)^{2}MA-GLVQ concept with Iterated Relevance Matrix Analysis (IRMA) [14,15] for bias mitigation. A domain-discriminative GMLVQ model $\mathcal{M}_D$ is trained using the domain sources as target labels. The relevance matrix Λ is often dominated by a single leading eigenvector, while additional class-relevant information may remain hidden in the orthogonal subspace of this eigenvector [16]. IRMA tackles this limitation by iteratively retraining GMLVQ under orthogonality constraints, removing previously identified dominant directions [15]. IRMA provides a more complete representation of underlying decision structure than relevance learning. This is particularly beneficial in scenarios with correlated or weakly relevant features, where multiple representations can support comparable predictive performance. Through k IRMA iterations, the accumulated domain-discriminative directions v_i are collected to form an orthonormal subspace basis $V_D = [v_1, \ldots, v_k]^\top \in \mathbb{R}^{k \times n}$ that captures the domain-dependent structure. Finally, the null-space projection $P_0 = I - V_D V_D^\top$ can be used in the transfer learning scenario to reduce or eliminate the domain information in/from the data [14]. Afterwards, the domain-mitigated $\mathcal{X}'_S = P_0 \mathcal{X}_S$ can be used to train a respective (interpretable) classification/regression machine learning model $\mathcal{M}'_S$ for the primary task, which is applicable to all mitigated data representations $\mathcal{X}'$. The implementation of this approach is shown in Fig. 1.

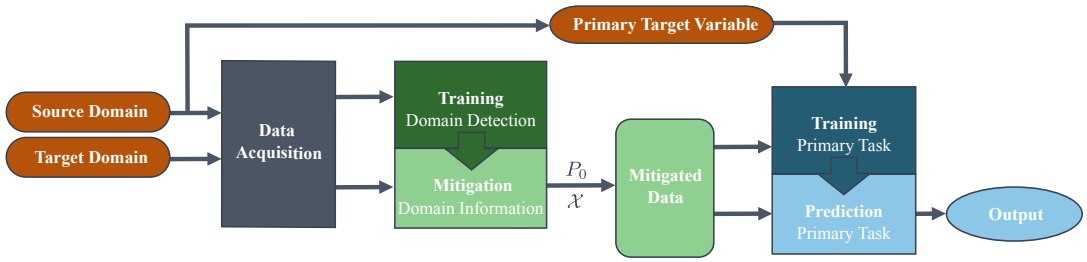


Fig. 1: Generalized workflow - transfer learning domain invariance. Adapted from [17]

3 Results

3.1 Application 1: Fully-Monitored to Restricted-Environment Transfer in Gait Phase Recognition

To evaluate the operationalization of the RAISSE methodology, we first present the results of transferring a gait phase recognition model across distinct measurement environments - specifically addressing a high-to-low sensor reduction from fully instrumented laboratory setups to lower monitored environments, as detailed in [17].

Problem Description & Experimental Setup: In biomedical movement analysis, obtaining high-precision ground truth labels for gait phases typically requires a Force Measurement Plate System (FMPS). However, FMPS-based setups are constrained by physical boundaries (e.g., only two complete steps per trial). Conversely, treadmill setups allow continuous, long-duration recordings but lack direct force-based labels. The goal of this application is to transfer gait-phase labels learned from a high-monitored source scenario A (FMPS, 10 participants) to an unlabeled, low-monitored target scenario B (treadmill, 8 participants). In both setups, kinematic motion was recorded using identical marker-based optical tracking and processed with the biomechanical model alaska/Dynamicus [18], resulting in a feature representation of $x \in \mathbb{R}^{120}$ (derived from joint angles and angular velocities). This setup involves a significant change in the experimental conditions and affects the human motion. The participants in Scenario B were a subset of those in Scenario A. Evaluation used subject-wise cross-validation with disjoint training and validation subjects, reducing the risk that domain discrimination is driven by subject-specific characteristics.

RAISSE Paradigm: Domain Mitigation Following the RAISSE paradigm, the environmental discrepancy between the two setups is modelled as a structured bias: A GMLVQ domain-classifier $\mathcal{M}_D$ was trained to discriminate between data from scenario A ($\mathcal{X}_A$) and B ($\mathcal{X}_B$). The model achieved $100\% \pm 0.0\%$ accuracy, indicating that the acquisition

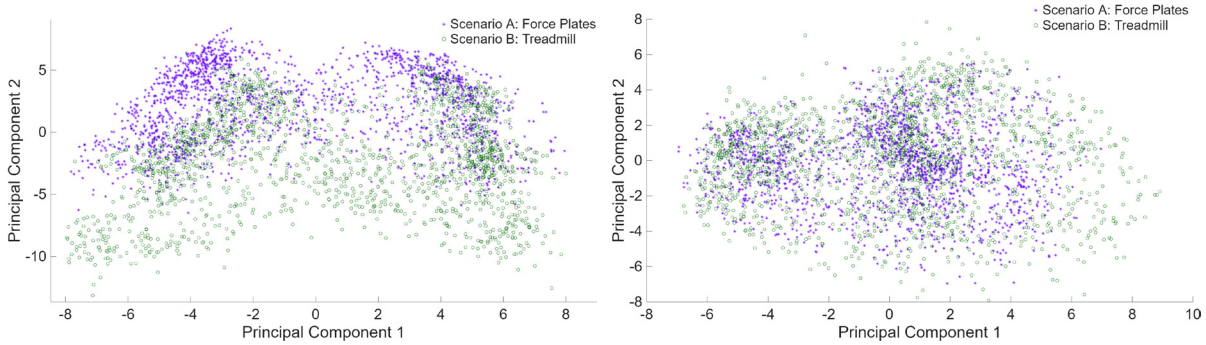


Fig. 2: PCA of the original data (left) and mitigated data (right) for both scenarios

setup induces a strong, systematic domain dependency in the feature space. To mitigate this dependency IRMA was applied to the source relevance matrix Λ_D with 25 iterations. Projecting the data onto P_0 successfully suppressed the environment-dependent structures, rendering the two scenarios substantially reducing their domain separability, i.e. $\mathcal{X}' = P_0\mathcal{X}$. The primary gait-phase classifier $\mathcal{M}'_A$ was then trained exclusively on the mitigated source data $\mathcal{X}'_A$ using the seven FMPS-derived gait-phase labels ($\mathcal{C}_A$).

Experimental Results: The evaluation demonstrates that domain mitigation preserves primary task accuracy while reducing setup-induced dependencies. On the unmitigated source data $\mathcal{X}_A$, the baseline GMLVQ primary classifiers achieved high gait-phase recognition accuracies of $94.7 \pm 0.3\%$ for the left foot ($\mathcal{M}_{A,l}$) and $94.9 \pm 0.3\%$ for the right foot ($\mathcal{M}_{A,r}$). Following the null-space projection, performance on the sanitized source data $\mathcal{X}'_A$ remained robust, yielding $93.4 \pm 0.3\%$ ($\mathcal{M}'_{A,l}$) and $92.6 \pm 1.3\%$ ($\mathcal{M}'_{A,r}$). Importantly, confusion matrix analysis revealed the remaining misclassifications occurred almost exclusively between temporally adjacent gait phases, suggesting that the underlying biomechanical transition structure is largely preserved in the feature space. When deploying the trained primary classifier $\mathcal{M}'_A$ directly onto the mitigated treadmill data $\mathcal{X}'_B$, visual plausibility checks conducted via alaska/Dynamicus [18] indicated a biomechanically plausible, continuous gait cycle reconstruction with natural temporal phase durations. As illustrated in Fig. 2, the original domain separation (left) is substantially reduced in the projected representation (right), supporting the central RAISSE hypothesis: removing setup-specific biases directly from the data representation supports transfer to restricted target environments.

3.2 Application 2: Measurement Setup-Shift Mitigation in Biogas Substrate Spectroscopy

To demonstrate the framework's capability in resolving subtle, low-magnitude measurement setup shifts, we evaluate RAISSE on a process spectroscopy scenario involving Near-Infrared (NIR) substrate analysis [19].

Problem Description & Experimental Setup: In industrial quality control and process monitoring, NIR spectroscopy is commonly employed to estimate key chemical properties of organic feedstocks, such as determining nitrogen content in biogas plant inputs. Spectral measurements were recorded in the wavelength range of 950 nm to 1650 nm at 2 nm resolution, resulting in an initial feature representation of 351 dimensions. To reduce dimensionality and extract dominant spectral characteristics, low-pass filtering was used. This yields a smaller, 30-dimensional feature space. In practical industrial operations, physical measurement rigs undergo measurement setup modifications, upgrades, or component replacements over time. In this study, data were collected across three distinct measurement setups(*init*, *update1*, and *update2*). Although these setups are based on the same fundamental measurement principles, slight physical differences between the individual setups can lead to systematic range-dependent effects, which can impair the predictive accuracy of the calibration models. We demonstrate the RAISSE workflow for classification tasks aimed at predicting high or low nitrogen concentrations under varying measurement conditions.

RAISSE Paradigm: Domain Mitigation Following the RAISSE workflow, the GMLVQ domain classifier $\mathcal{M}_D$ was trained on the original, unmitigated, low-pass filtered spectra to discriminate the nitrogen concentration using the data of the *init* measurement setup. Then the obtained model is applied on the data of the measurement setup *update1* and *update2*. In the next step a model is trained to distinguish between the three measurement setups. The model achieved a balanced accuracy of 64.9%, confirming the presence of a subtle yet systematic setup dependency. To neutralize this setup-induced bias, IRMA was applied with $k = 5$ iterations. The resulting $V_D \in \mathbb{R}^{5\times 30}$ was used

to construct the null-space projection matrix P_0, which was subsequently applied to clean the dataset.

Experimental Results: It can be shown that low and high nitrogen concentrations can be distinguished using the spectral data under the init setup, where GMLVQ with three prototypes per class achieves an accuracy of $92.6 \pm 0.08\%$ However, when this classifier is applied to the two other measurement setups, its performance drops to $78.6 \pm 19.2\%$, indicating a setup-induced domain shift. This behaviour is also reflected in Fig. 3, which visualises the mean spectra of the respective classes and their controls: one set shows the original low-pass-filtered data, while the other depicts the corrected projected data. Using FA(IR)^{2}MA, this drift can be eliminated, and the resulting data no longer allow the measurement scenarios to be distinguished. Yet, when attempting to solve the original classification task on the cleaned data, the performance drops to $86.9 \pm 2.7\%$, demonstrating that the drift overlaps with the spectral information relevant for nitrogen concentration. Consequently, a separate model must be trained for each new measurement setup, since a single model cannot provide reliable predictions when each modification of the experimental setup unpredictably alters the signal properties used to measure physical quantities. This finding illustrates that the RAISSE principle is essential for obtaining trustworthy models predictions.

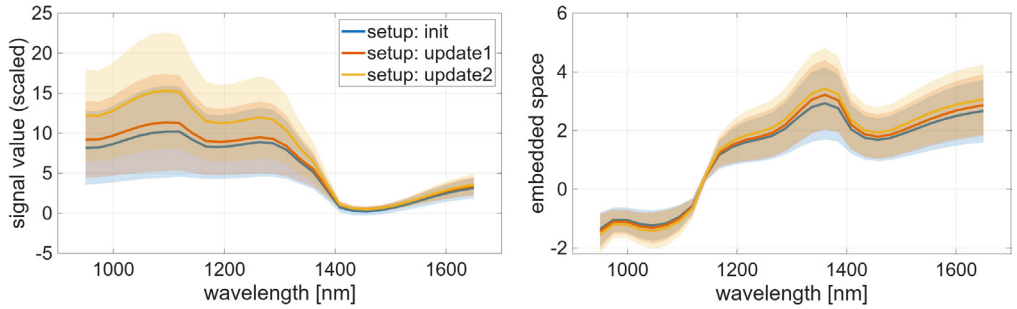


Fig. 3: Mean spectra with dispersion of the three measurement setups of the original low-passed spectra (left) and the resulted mitigated spectra after IRMA is applied (right)

4 Discussion and Conclusion

This work illustrates the RAISSE methodology as an interpretable and mathematically grounded framework for tackling transferability and domain discrepancies in restricted environments. The RAISSE principle fundamentally shifts the paradigm by treating transferability as a representation problem rather than a model adjustment problem. The core principle of RAISSE expands upon the foundational concepts of prototype-based relevance learning. While relevance learning with GRLVQ was designed to isolate compact, task-relevant feature subsets and sensor reduction in homogeneous datasets, unmitigated domain shifts pose a fundamental threat: metric adaptation risks assigning high relevance to setup-dependent artifacts rather than true physical characteristics. RAISSE addresses this limitation by using FA(IR)^{2}MA-GLVQ to explicitly detect and quantify domain-specific biases directly in the data representation, mitigating these biases via null-space projection.

The presented applications demonstrate the broad applicability of this workflow across distinct modalities, task types, and bias magnitudes. In biomechanics, RAISSE enabled a biomechanically plausible transfer of gait-phase recognition from a highly instrumented laboratory setup to an unlabelled, sensor-reduced treadmill scenario. Despite suppressing environment-dependent kinematic structures, high classification accuracy was maintained, with misclassification restricted almost exclusively to temporally adjacent gait phases, indicating that substantial task-relevant biomechanical structure was retained. In spectroscopy and industrial process control, the application to NIR spectroscopy demonstrated the framework's capability to detect the presence of drift and the fact that the drift and the information about the low/high nitrogen concentration is overlapping. The RAISSE principle demonstrates this, and appropriate measures can be derived from it to ensure robust and reliable classification models.

Crucially, these applications highlight a fundamental insight regarding the architectural flexibility and origin of the framework: while the initial bias detection and mitigation phase relies on interpretable prototype models like GMLVQ, the choice of the primary task processing can be tailored entirely to the target application. Whether deploying classical GRLVQ relevance learning to further prune redundant sensors/spectral channels, employing prototype-

based classifiers for time-series recognition, or utilizing regression algorithms for continuous property estimation, the framework offers a flexible, task-agnostic mechanism to reduce domain-discriminative variance while retaining task-relevant information as far as possible.

Despite these advantages, certain limitations of the current framework must be acknowledged. While the null-space projection effectively suppresses domain-discriminative signatures, each mitigation step inherently alters the geometry of the representation space. As a consequence, removing domain-separating directions carries the risk of inadvertently discarding task-relevant variance required for the primary predictive goal. Balancing bias mitigation with the preservation of task-aware information constitutes a non-trivial trade-off. Consequently, systematically optimizing this trade-off through task-aware stopping criteria and guided mitigation strategies remains a key subject of future investigation.

Overall, the results support the central RAISSE hypothesis: reliable transfer learning, model generalization, and legacy data sanitization across restricted environments can be achieved through systematic mitigation of domain-dependent structures within the data representation itself. By avoiding complex target-side fine-tuning or non-transparent domain alignment networks, RAISSE provides high practical value for smart system engineering in domains where explainability, resource efficiency, and safety are paramount. Promising application fields include industrial condition monitoring across machinery with varying operational hours, simulation-to-real transfer, and quality inspection scenarios where physical sensors or measurement rigs are modified, upgraded, or relocated over time.

References

[1] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, 2010.
[2] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, 2018.
[3] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2130, 2016.
[4] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, 2019.
[5] M. Riedel, F. Rossi, M. Kästner, and T. Villmann, "Regularization in Relevance Learning Vector Quantization Using l1-Norms," *Comput. Intell.*, 2013.
[6] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1521–1528, 2011.
[7] T. Kohonen, "Learning Vector Quantization," *Neural Networks*, vol. 1, p. 303, 1988.
[8] A. Sato and K. Yamada, "Generalized Learning Vector Quantization," in *Advances in Neural Information Processing Systems*, vol. 8, MIT Press, 1995.
[9] B. Hammer, M. Strickert, and T. Villmann, "On the Generalization Ability of GRLVQ Networks," *Neural Process. Lett.*, vol. 21, no. 2, pp. 109–120, 2005.
[10] P. Schneider, M. Biehl, and B. Hammer, "Distance learning in discriminative vector quantization," *Neural Comput.*, vol. 21, no. 10, pp. 2942–2969, 2009.
[11] M. Biehl, B. Hammer, and T. Villmann, "Prototype-based models in machine learning," *WIREs Cogn. Sci.*, vol. 7, no. 2, pp. 92–111, 2016.
[12] S. Saralajew, L. Holdijk, M. Rees, and T. Villmann, "Robustness of Generalized Learning Vector Quantization Models Against Adversarial Attacks," in *Advances in Self-Organizing Maps, Learning Vector Quantization, Clustering and Data Visualization*, pp. 189–199, Springer, 2020.
[13] K. Bunte, P. Schneider, B. Hammer, F.-M. Schleif, T. Villmann, and M. Biehl, "Limited Rank Matrix Learning, discriminative dimension reduction and visualization," *Neural Netw.*, vol. 26, pp. 159–173, 2012.
[14] M. Kaden *et al.*, "FA(IR)2MA-GLVQ – A hidden-feature-bias mitigation approach for fairness in classification learning based on generalized matrix learning vector quantization," *Neurocomputing*, vol. 678, p. 133200, 2026.
[15] S. Lövdal, E. L. van den Brandhof, and M. Biehl, "Iterated relevance matrix analysis for improved classification and robustness in prototype-based learning schemes," *Neurocomputing*, vol. 670, p. 132474, 2026.
[16] M. Biehl, B. Hammer, F.-M. Schleif, P. Schneider, and T. Villmann, "Stationarity of Matrix Relevance LVQ," in *Int. Joint Conf. Neural Netw. (IJCNN)*, pp. 1–8, 2015.
[17] T. Pfaff, D. Möbius, T. Villmann, and M. Kaden, "Gait-Phase Classification using Environmental Dependent Measurements and Transfer Learning," in *IEEE MetroXRAINE*, 2026.
[18] H. Hermsdorf, N. Hofmann, and A. Keil, "Bewertung digital erfasster Bewegungen mit Dynamicus," in *Homo Sapiens Digitalis - Virtuelle Ergonomie und digitale Menschmodelle*, pp. 329–336, Springer, 2016.
[19] R. Schubert *et al.*, "A White-Box Workflow for the Prediction of Food Content From Near-Infrared Data Based on Fourier-Transformation," in *13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–5, 2023.

A data-driven optimization approach towards better automation in industrial fused granular fabrication

Nikolai Kirchner, Lukas Tuchscherer, Celina Hennig, Johannes Abicht, Steffen Ihlenfeldt
Fraunhofer-Institute for Machine Tools and Forming Technology IWU, Pforzheimer Str. 7a, 01189 Dresden, nikolai.kirchner@iwu.fraunhofer.de

Abstract: Since its introduction additive manufacturing has revolutionised small scale manufacturing and prototyping, due to its high production speed, low cost and freedom of design in component modelling. Concerning the additive manufacturing of thermoplastics, Fused Granular Fabrication (FGF) offers even greater benefits regarding production speed and cost compared to Fused Filament Fabrication (FFF), showcasing great potential for industrial application and manufacturing of large components. Yet in contrast to readily available FFF printers, many FGF printers lack material and process datasets, resulting in a high amount of manual planning and test-prints by skilled machine technicians. The technicians must rely on their experience, carefully plan the toolpath and manually adjust the gcode with the necessary process parameters, negating the processes advantages in production speed and low material waste. Currently, this is proving to be a major hindrance for the wider industrial application of FGF-technology. To address this issue, a calibration part is developed with which a multitude of process parameters are simultaneously printed. Key regions of interest of the calibration part are measured and a database is established, combining process parameters with the measured geometric values. A process-optimization tool is established, embedded into a postprocessor and integrated into an open-source slicing software. By automatically optimizing the process parameters for the generated toolpath, the manufacturing time and the waste material could be significantly reduced to achieve the desired geometric features.

Keywords: additive manufacturing, fused granular fabrication, process optimization

1 Introduction

Due to its freedom of design, low cost and high production speed, additive manufacturing has seen a steady growth in application. With a strong increase of 3D-printing applications in the industrial sector, the focus has largely remained on small scale 3D-printers, due to the challenges, such as prolonged printing time, posed by large scale models and printers [1]. Fused granular fabrication (FGF) tackles this issue by processing pellets instead of filament, rapidly increasing the production speed while lowering production costs, compared to the widely established fused filament fabrication (FFF) process. While FGF allows for larger volumes to be printed, it faces challenges considering the accuracy of the prints [2], [3]. The absence of extrusion dynamic modelling in FGF [4] and the inherent variability of pellet-based extrusion systems, compared to filament-based systems, have limited process standardization, reducing the predictability and quality of FGF manufacturing. Even inconsistent extrusions, leading to geometric deviations of the print due to varying bead widths, remain a challenge in FGF manufacturing [5]. This necessitates the need for additional preparation time to achieve high-quality prints [2].

To increase dimensional accuracy in 3D-printing, cubical structures have been printed as calibration parts with varying process parameters, to identify their effect on the print quality [6], [7]. However, a more fundamental understanding of the influence of process parameters can be obtained by analysing individual bead widths, as stable and predictable beads form the basis for the overall geometry and quality of FGF prints. If the relation between the bead widths and process parameters can be correctly modelled, their correct prediction would greatly reduce the amount of planning time and waste material necessary for high-quality prints. By analysing single bead widths, M. Layher et al. modelled the correlation between bead geometry, feed rate, layer height and spindle revolutions in the FGF process and developed a software application, with which the process parameters can be calculated for the desired bead geometry [8]. M. Hakimiy and K. Drödel conceptualized an automated parameterisation approach of

the FGF-process, by integrating a profile sensor into a custom FGF printer, varying printing parameters and measuring geometric accuracy, such as the bead width [9].
By applying similar methods, this work presents an approach for establishing and integrating a material specific process database for an industrial FGF printer, to reduce the amount of preparation time and waste material for high-quality prints. Favourable parameter combinations need to be identified, as shown in Fig. 1. A calibration part is developed, with which a multitude of parameter combinations can be printed in one print, increasing the efficiency and speed with which the database is established. In accordance with the state of the art, the bead width of the calibration part is analysed. The data is then processed and integrated into the planning process via gcode to predict favourable parameters.

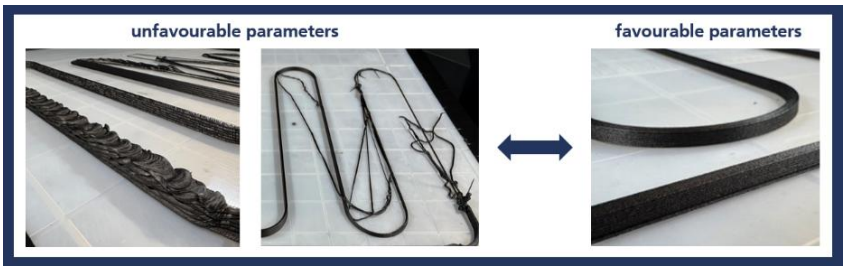


Fig. 1. Varying quality of FGF-3D-printed parts, resulting from different combinations of process parameters.

2 Methodology

The presented methodological approach is split into two steps, as presented in Fig. 2. Firstly, a calibration part is developed in section 2.1, to measure the effect of different process parameters on the 3D-printed geometric features. With the results a database is developed, matching the process parameters with the resulting bead widths. Following the implementation of the database, a process optimizer is developed in section 2.2, automatically integrating the calculated parameters into the generated toolpath via gcode.

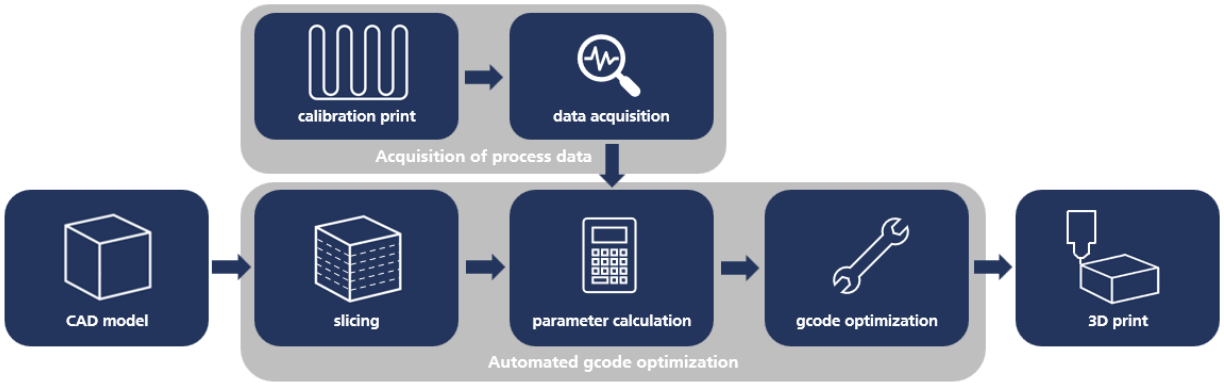
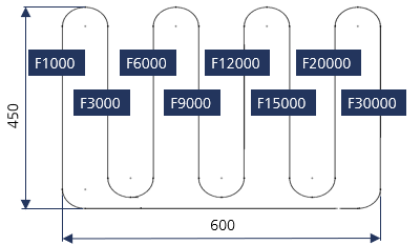


Fig. 2. Process flow of the presented methodology. The approach is split into two sections, namely the acquisition of process data and the automated gcode optimization.

2.1 Acquisition of process data

Different parameter combinations, such as printing speed, extrusion rate and temperature profile of the extruder, directly affect the bead width and its quality. Most FGF printers are equipped with a spindle in the print head. Thus, apart from the material in use and temperature profile of the extruder, the extrusion rate is primarily determined by the spindle revolutions. As the ideal temperature profile of the material is usually indicated in the material datasheet, the primary process parameters that are analysed for the bead widths are the travel speed and spindle revolution in dependence of the set layer height and nozzle diameter.

The calibration part is modelled into clear sections for the measurement of varying bead widths, as shown in Fig. 3. These sections are deliberately modelled as straight lines, to minimize the effects of acceleration and deceleration of the printer kinematics on the bead width for a reliable correlation between travel speed of the printer and extrusion rate. Each section is assigned a fixed printing speed through manual implementation in the gcode across all layers, resulting in the different parameter combinations in one print, accelerating the rate at which data can be obtained. Each calibration part is printed with a specified nozzle diameter, layer height and spindle revolution. The different parameters with which the part is printed are listed in Fig. 3.
The method is validated with the SEAMHex2, as demonstrated in Fig. 4, a FGF printer for industrial application with an output rate up to 10 kg/h and a printing speed up to 1 m/s [3]. The prints are done with carbon-fibre-reinforced polyamide 6 (PA6-CF40).
Per section, 5 measurements of the bead width are done with a calliper. The measurements are averaged and the combination of process parameters and bead widths are saved in a xlsx-file.



Parameters	Values							
nozzle diameter [mm]	1	2	3	4				
layer height [mm]	0.2	0.5	1.0	1.5	2.0			
printing speed [mm/min]	1000	3000	6000	9000	12000	15000	20000	30000
spindle revolutions [revs/min]	5	10	20	30	40	50	…	150

Fig. 3. The design of the calibration part with assigned printing speeds, noted as F-commands, and the parameters with which the calibration parts were printed. While the printing speeds are varied in a single calibration part, different nozzle diameters, layer heights and spindle revolutions require the print of a new calibration part.

2.2 Automated gcode optimization

To integrate the acquired process information from the established database into the slicing process, a postprocessor with an embedded process-optimizer is developed for OrcaSlicer, an open-source slicer. The postprocessor overwrites the necessary M-commands, as a specialised 3D-printer is used, before the process-optimizer is initiated. Depending on the material, for the mechanical stability of the printed part, a layer time must be adhered. The optimizer determines the travel distance and required travel speed to achieve the layer time. As the layer time isn't a specific, but rather a region of values, the closest travel speed in the database is chosen.
Manual bead widths are set in the slicing software. After slicing, the optimizer analyses the gcode, identifies the desired bead width and in combination with the calculated travel speed, interpolates the necessary spindle revolution. This operation is carried out for each layer and depending on the model's geometry, different printing speeds and spindle revolutions can be calculated per layer. These values are saved as variables at the beginning of each layer within the gcode, eliminating further need for manual parametrization.

3 Results

3.1 Acquisition of process data

256 calibration parts were printed with PA6-CF40 to cover the different spindle revolutions, nozzle diameters and layer heights. An approximate linear relation between bead width and spindle revolutions could be identified for the set printing speeds and for layer heights ≥ 0.5 mm, as demonstrated by the results of the 2 mm nozzle in Fig. 4. All nozzle diameters showcased a linear relation, while differing in slope and intercept. Outliers within the approximated linear curves can be attributed to inaccuracies in measurement. Only the top layers of the calibration part were measured, as certain parameter combinations and bead widths resulted in elephant's foot of the bottom layers, a common effect in 3D printing, that needs to be analysed in future works. As to be expected, depending on

the nozzle diameter and layer height, not all spindle revolutions and printing speeds were compatible for a high-quality print, with some resulting in poor layer adhesion or over-extrusion, resulting in poor surface quality and highly variable bead widths. These parameter combinations were labelled as inadequate combinations within the database and not used for further interpolation. This explains the comparatively low number of measurements for the printing speeds of 1000 and 30000 mm/min in Fig. 4, as most printed parts showcased inadequate surface quality for these parameters.

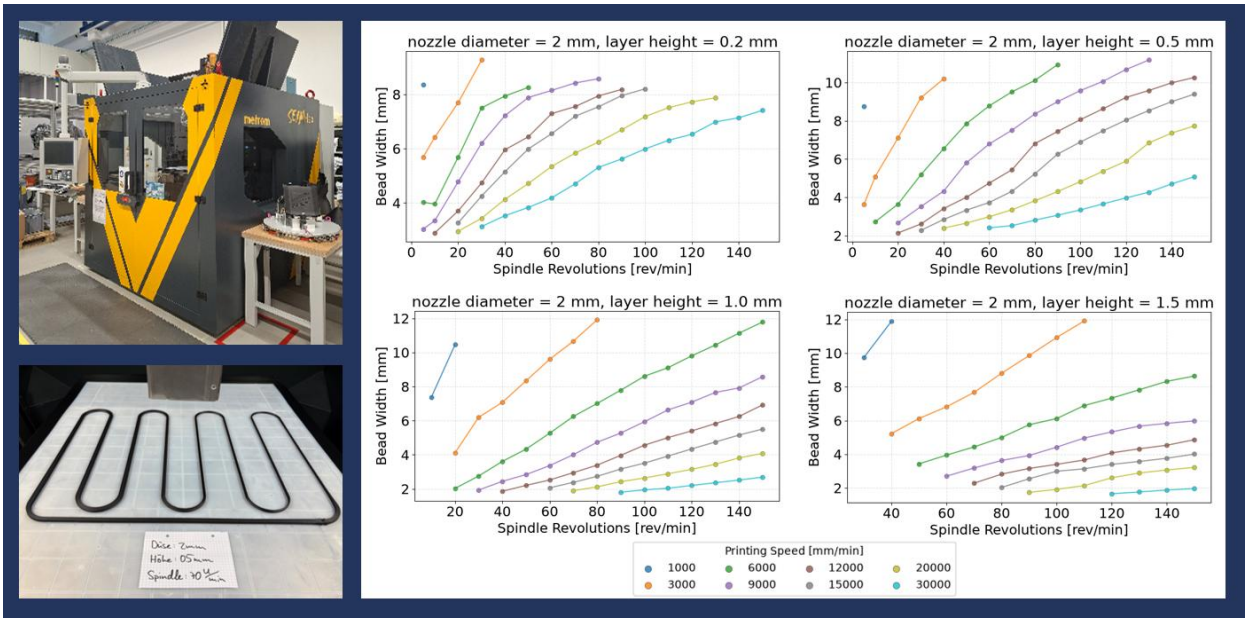


Fig. 4. Upper left: The SEAMHex2 FGF printer. Bottom left: A printed calibration part. Right: Combined measurements of the bead widths for the calibration parts printed with a 2 mm nozzle. Every graph combines the measurements of 16 calibration parts for the specified layer height.

3.2 Automated gcode optimization

To test the effectiveness of the optimizing tool on the manufacturing time and waste material, four workpieces were developed and sliced, as seen in Fig. 5. For experimental validation, the four workpieces were each printed with, and without the use of the optimizer. For each workpiece, the order of the print with or without optimizer was varied. As every workpiece is printed twice and to eliminate the possibility of the participant memorizing the parameters, the layer height was varied between 0.5 and 1 mm for prints. Manual adjustments to the gcode were allowed for both test cases and every workpiece was printed as many times as needed, until the line width of the workpiece had a maximum deviation of 10 % of the set target bead width. 10 points were measured by calliper per workpiece and the values averaged. For workpieces with two walls, the allowed deviation was 10 % of two bead widths. The planning and printing time was measured for each workpiece, and the waste material was weighed. All bypass material and print attempts apart from the final workpiece were considered as waste material. As presented in Fig. 6, the results of the experiment show that while both manufacturing time and waste material vary greatly between workpieces, both factors could be significantly reduced with the optimizer, due to a lower number of necessary test-prints necessary to find the ideal parameter set. To correct the systematic deviation in printing time between prints with 0.5 and 1 mm layer heights, the printing time for prints with a 0.5 mm layer height were scaled by the factor 2 (Fig. 6a). In this case study, manufacturing times could be reduced by 30 – 77 % and waste material could be reduced by 59 – 99 % with the optimizing tool.

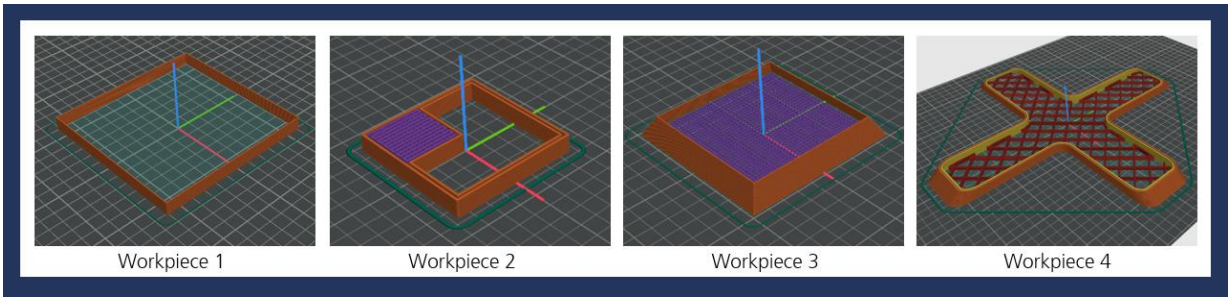


Fig. 5. Developed workpieces for experimental FGF print and validation of the optimizing tool. The workpieces resemble different levels of complexity in planning, through the addition of infills, overhangs and double walls.

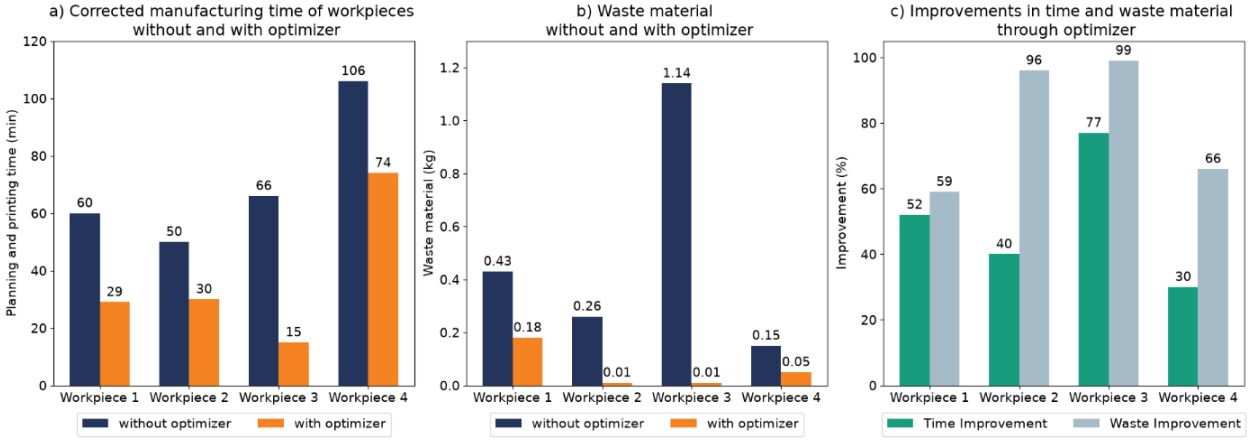


Fig. 6. Results of the experimental validation of the optimizing tool. (a) The corrected manufacturing time for the FGF-print of the workpieces. (b) Weighed waste material until completion of target criterion for the print. (c) Comparative manufacturing time and waste improvements (reduction) with the optimizing tool.

4 Discussion and Outlook

Based on the data collected from the printed calibration parts, the presented approach proved effective in significantly reducing planning time and waste material in the case study. For future works a validation with multiple participants is necessary. A wide variation in planning time and waste material reduction is expected, resulting from the participants differing experience levels with the FGF-process and printer. In accordance with the results of this case study however, a clear trend towards planning time and waste material reduction is expected with the process optimizer. Especially in environments where the FGF printer is operated without dedicated process specialists or when testing new materials, the approach is expected to provide the greatest value. While verified on the SEAMHex2 printer, the approach can be applied to various FGF printers with spindle extrusion systems by modifying the starting parameters and postprocessor. However, further studies are needed to verify the method for different FGF printers. To widen the application of the process optimizer, the described methodology needs to be applied to additional materials to build further databases. Since a linear relation of the variables could be determined for PA6-CF40, future data acquisition can greatly reduce the number of calibration prints for data acquisition, as merely a different slope and starting point of the curve is expected for different materials, if linear relations can be identified. In this case, in theory, the amount of necessary calibration prints can be reduced by up to 87.5 % for future materials.

For further automation of data acquisition of new materials, a Wenglor MLWL132 laser line sensor was integrated into the SEAMHex2 printer, for automated measurement of bead widths of the calibration part, greatly increasing the number of measurements while reducing manual labour. 8 calibration parts were measured as a proof of concept, with bead widths varying between 1,97 – 10,69 mm. The bead widths could be correctly identified with a tolerance of 0,2 mm in 90,625 % of all cases. By adding another profile sensor into the setup, a greater accuracy is expected. By further adding an encoder into the setup, a complete digital reconstruction of the printed part could be achieved, combining the data-logged machining parameters with the print results. For future prediction of

parameter combinations for nonlinear materials, the application of machine learning models should be tested, to take the wider array of parameters, such as extruder temperature settings and nozzle pressures, into account.

5 Conclusion

In this paper, a methodology is presented with which the efficiency of industrial 3D-printers can be increased by integrating process data into the workflow, to greatly reduce manufacturing time and waste material. First, a calibration part was developed to sequentially print varying parameter combinations in one print. Multiple calibration parts were printed with PA6-CF40 on the SEAMHex2 printer. The bead widths of the calibration parts were measured and saved in a database along with the process parameters. To automatically integrate the process parameters into the slicing and planning process, an optimization tool was developed and integrated into a postprocessor for an open-source slicer. In a case study, both manufacturing time and waste material could be significantly reduced by automatically integrating the process data into the planning process.

Acknowledgements

With the open-topic and cross-industry funding program "Industrielle Gemeinschaftsforschung" (IGF), the Federal Ministry for Economic Affairs and Energy supports pre-competitive research projects. The presented work is part of the IGF project "Automatisierte Prozesskette: Vom 3D-Scan großer Teile zur Additiven Fertigung (Scan2MEX)" (01IF2444N) and is funded by the Federal Ministry for Economic Affairs and Energy based on a decision by the German Bundestag. The authors gratefully acknowledge the financial support that made this work possible.

References

[1] J. Shah, B. Snider, T. Clarke, S. Kozutsky, M. Lacki, and A. Hosseini, 'Large-scale 3D printers for additive manufacturing: design considerations and challenges', *Int. J. Adv. Manuf. Technol.*, vol. 104, no. 9, pp. 3679–3693, Oct. 2019, doi: 10.1007/s00170-019-04074-6.

[2] J. Kalle, K. Joni, S. Alexander, and O. Juhani, 'Potential and Challenges of Fused Granular Fabrication in Patternmaking', *Int. J. Met.*, vol. 17, no. 4, pp. 2469–2476, Oct. 2023, doi: 10.1007/s40962-023-00989-9.

[3] S. Ihlenfeldt, W.-G. Drossel, M. Kausch, C. Friedrich, T. Wiese, and L. Jankowsky, 'SEAMHex - Fast 6D Additive Manufacturing using an innovative Screw Extruder applied on a Hexapod Parallel Kinematic', *Technol. Lightweight Struct. TLS*, vol. 5, no. 1, pp. 1–14, 2021, doi: 10.21935/tls.v5i1.166.

[4] A. Pricci, M. D. de Tullio, and G. Percoco, 'Analytical and Numerical Models of Thermoplastics: A Review Aimed to Pellet Extrusion-Based Additive Manufacturing', *Polymers*, vol. 13, no. 18, p. 3160, Jan. 2021, doi: 10.3390/polym13183160.

[5] G. D. Goh, K. K. Wong, N. Tan, H. L. Seet, and M. L. S. Nai, 'Large-format additive manufacturing of polymers: a review of fabrication processes, materials, and design', *Virtual Phys. Prototyp.*, vol. 19, no. 1, p. e2336160, Dec. 2024, doi: 10.1080/17452759.2024.2336160.

[6] L. Santana, J. Lino Alves, and A. da Costa Sabino Netto, 'A study of parametric calibration for low cost 3D printing: Seeking improvement in dimensional quality', *Mater. Des.*, vol. 135, pp. 159–172, Dec. 2017, doi: 10.1016/j.matdes.2017.09.020.

[7] S. Stopp, T. Wolff, F. Irlinger, and T. Lueth, 'A new method for printer calibration and contour accuracy manufacturing with 3D-print technology', *Rapid Prototyp. J.*, vol. 14, no. 3, pp. 167–172, May 2008, doi: 10.1108/13552540810878030.

[8] M. Layher, L. Eckhardt, A. Hopf, and J. Bliedtner, 'Development of a Process Model for Bead Deposition Rates and Cooling Behavior of Large Scale Additive Manufacturing Parts', in *Industrializing Additive Manufacturing*, M. Meboldt and C. Klahn, Eds., Cham: Springer International Publishing, 2021, pp. 223–240. doi: 10.1007/978-3-030-54334-1_16.

[9] M. Hakimiy and K. Dröder, 'A concept for in-situ parameterisation of a fused granular fabrication process: A material independent approach', *Procedia CIRP*, vol. 138, pp. 573–578, Jan. 2026, doi: 10.1016/j.procir.2026.01.099.

Dark Company: Autonomous Multi-Agent Systems and the Decoupling of Corporate Output from Headcount

Jens Kneifel[1], Robin Roj[1]
[1]Forschungsgemeinschaft Werkzeuge und Werkstoffe e. V.,
Papenberger Straße 49, 42859 Remscheid, kneifel@fgw.de, www.fgw.de

Abstract: Lights-out manufacturing has demonstrated that physical production can run without on-site human presence, yet the administrative, cognitive, and commercial functions of a manufacturing enterprise have remained bound to human labor. This contribution introduces and delimits the concept of the Dark Company: the organizational analog of the dark factory, in which an orchestrated network of autonomous, large-language-model-based agents extends beyond the shop floor to the entire enterprise. Using a mid-sized manufacturer of hand tools with metal processing as its core technology as a running example, a reference configuration is derived department by department, covering engineering design, manufacturing, intralogistics, maintenance, purchasing, sales, marketing, the support functions as well as the management. This is combined into a conceptual architecture and an exemplary process flow. The analysis indicates that corporate output can in principle be decoupled from headcount, but that the binding constraints migrate from labor capacity to orchestration, to the interface between agent-based office work and physical automation, to error containment and legal accountability. The contribution is conceptual: it consolidates a loosely used term, exposes the maturity gaps, and delimits the residual human role.

Keywords: Agentic AI; Artificial Intelligence; Dark Factory; Dark Office; Robotics;

1 Scientific Background and Motivation

The term 'Dark Company' denotes a convergence phenomenon at the intersection of automation, operational opacity, and regulatory withdrawal. In analogy to lights-out ("dark factory") manufacturing, which produces around the clock without on-site human presence [1], the concept of a Dark Company is a primarily digital enterprise whose administrative, cognitive, and value-creating processes are governed by an orchestrated network of autonomous AI agents [2]. These act not as passive assistance systems but as goal-directed decision-makers that conduct the operational business as virtual employees collaborating within a shared virtual office [3].

The economic driver is the decoupling of corporate output from headcount. In knowledge- and service-based organizations, personnel traditionally constitutes the largest cost block and ties the achievable output directly to the number of employees. Where this labor is substituted by scalable compute, the viability threshold of a business model falls, so that even previously prohibitive business fields become operable. Technologically, the lever shifts from rule-based robotic process automation (RPA) [4] to agentic process automation (APA), in which cognitive planning engines independently structure unstructured workflows, anticipate errors, and react dynamically to changing conditions. As the concept remains loosely defined, this contribution consolidates the term, proposes a reference framework, and examines the state, economic rationale, and limits of agent-run organizations.

The physical half of this proposition is comparatively well documented. Cyber-physical production systems and the smart factory have been described as the technological substrate of Industry 4.0, whereas the subsequent Industry 5.0 reintroduces human-centricity, sustainability, and resilience as explicit design objectives [5]. Lights-out factories, in which machining, handling, and inspection proceed without on-site personnel, have been reviewed systematically; the obstacles reported there are less the individual machine than reliability, the handling of unanticipated exceptions, and a residual dependence on skilled staff [1]. Within the shop floor itself, the delegation of control decisions to software agents is not new: agent-based distributed manufacturing control has been

investigated for decades, with holonic and heterarchical architectures proposed in place of centralized scheduling [6]. What these strands share is a scope restriction to the production system.

On the administrative side, robotic process automation established that mental work can be executed by software, but remained confined to deterministic, rule-based sequences on stable interfaces [4]. Agents built on large language models relax this restriction: surveys of their construction distinguish profile, memory, planning, and action modules and document applications across engineering and the sciences [2]. Multi-agent configurations add role assignment, communication protocols, and collective reasoning, and are reported to outperform monolithic models on decomposable, multi-step tasks while introducing new failure modes such as the propagation of errors between agents [3]. Function-specific evidence exists for several classical departments: deep generative models are established in engineering design for the synthesis and evaluation of geometry under performance constraints [7], and machine-learning-based predictive maintenance is among the most extensively reviewed Industry 4.0 applications [8]. Labor-economics research quantifies the substitution side, estimating employment and wage effects of industrial robot adoption [9]. The governance question, finally, has been treated separately: accountability for AI-based decisions is conceptualized through attributability, answerability, and auditability, all of which presuppose an identifiable human or institutional principal [10].

The literature therefore treats the dark factory and the agent-run office as separate objects and examines individual departments in isolation. What is missing is a consolidated organizational reading in which the physical and the cognitive automation layers close into a single loop, from customer inquiry through design, production, and delivery to invoicing. This paper addresses that gap. It defines the Dark Company as the organizational analog of the dark factory; it derives a department-by-department reference configuration for a concrete mid-sized hand-tool manufacturer, thereby making the interface between agent-based office work and physical automation explicit; and it tests the decoupling thesis against maturity, resilience, and accountability constraints. The contribution is conceptual rather than empirical: existing, individually demonstrated capabilities are composed into one organizational architecture whose limits are then delimited.

2 Materials and Methods

The analysis combines a synthesis of current literature and vendor practice with a conceptual reference architecture and an exemplary process flow. The general structure of such a system is shown in Fig. 1.

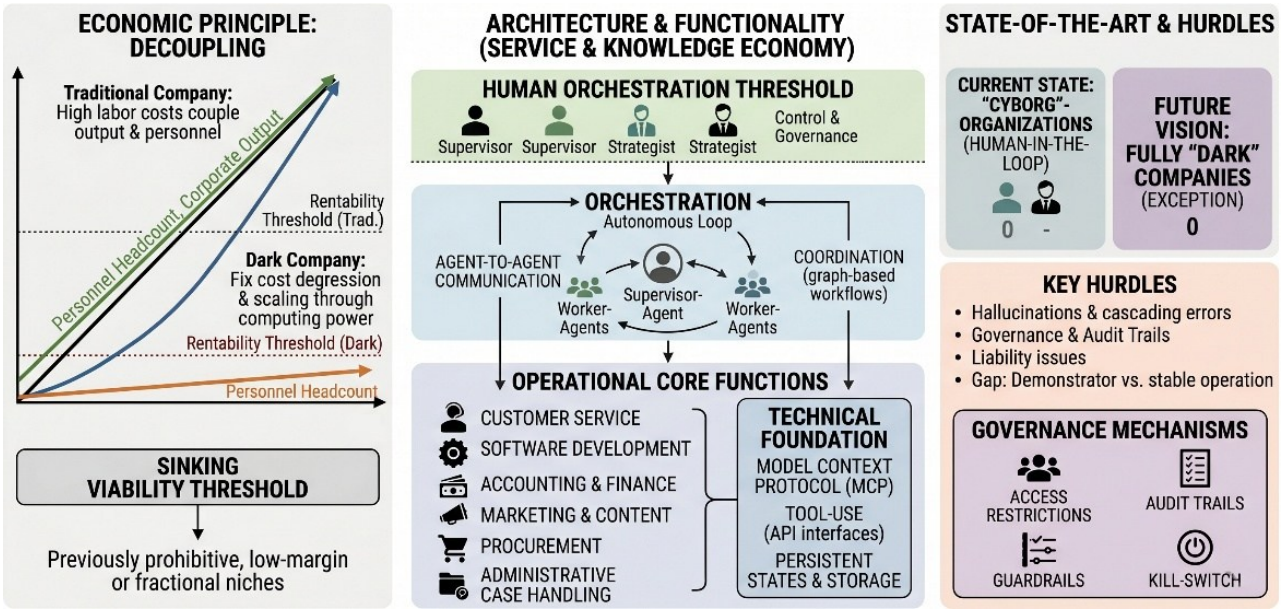


Fig. 1. Overview of the concept of a Dark Company

To keep the analysis specific, all subsequent considerations are grounded in a single reference case: a mid-sized industrial manufacturer of hand tools, such as pliers, wrenches, screwdrivers, and striking tools, with metal

processing as its core technology. The company operates a process chain of forging or cold forming, machining, heat treatment, surface finishing, and assembly; it sells a catalog of some thousand mature variants into industrial distribution; and it employs a few hundred people, of whom a substantial share work in administration, engineering, sales, and support rather than at the machine. Its products are geometrically stable, its process chain is well characterized, and its order patterns are largely repetitive, which makes it a favorable rather than a representative candidate for far-reaching automation. The reference configuration is developed along the classical functional departments of such a company: engineering and design, manufacturing, intralogistics and material handling, maintenance, purchasing, sales, marketing, the support functions of quality management, accounting, human resources, information technology and compliance, and finally management. For each department three questions are asked: which decisions are taken, which data and actuation interfaces an agent would require, and what remains for the human. The result is a target state that is plausible on the basis of individually demonstrated capabilities, not an implemented system. The criteria applied to that are technical maturity, the completeness of the interface between office and shop floor, resilience under fault conditions, and legal attributability.

3 Results

The operating model is illustrated in Fig. 2 by a representative task: the autonomous handling of a customer inquiry.

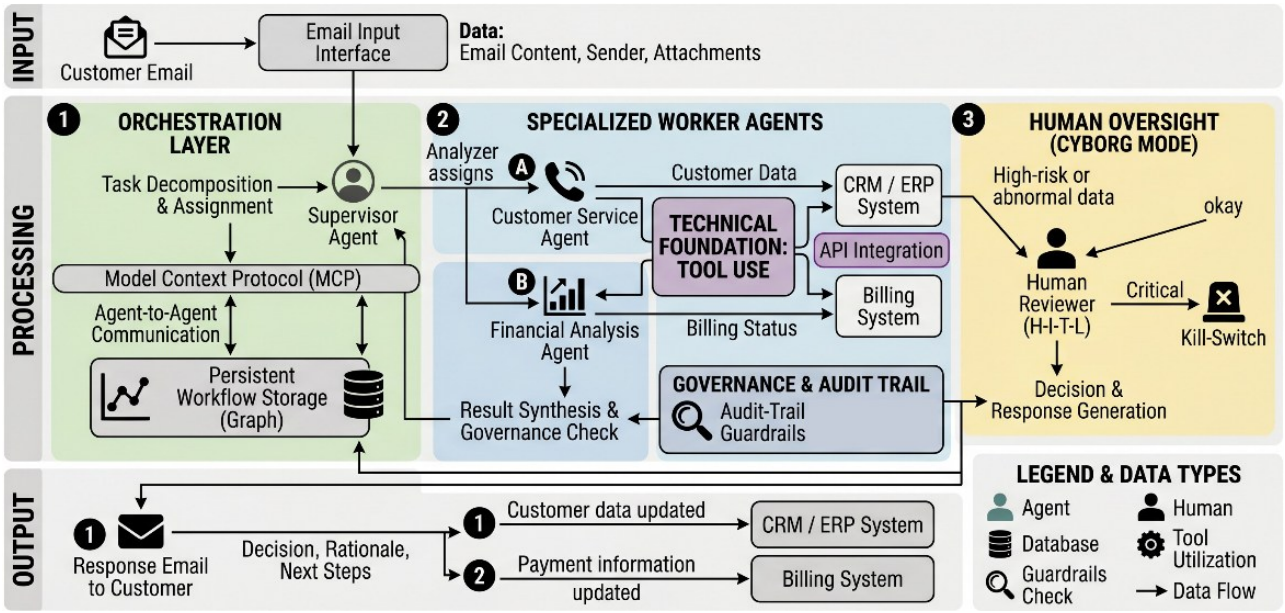


Fig. 2. Typical data flow in a Dark Company illustrated by an autonomous customer-inquiry workflow

An incoming e-mail is captured and passed to the supervisor agent, which decomposes and assigns the task. Specialized worker agents – here a customer-service and a financial-analysis agent – query CRM/ERP (Customer Relationship Management, Enterprise Resource Planning) and billing systems via tool-use through their APIs (Application Programming Interface); every step is logged in the persistent, graph-based workflow memory. The results are synthesized and subjected to a governance check; high-risk or ambiguous cases escalate to a human controller (H-I-T-L; human-in-the-loop), who can approve, reject, or trigger the kill-switch. Only thereafter are the response e-mail and the internal records generated or updated and logged in an audit-proof manner.

Engineering and design is the department in which agent-based work is most directly continuous with established practice. For a hand-tool portfolio the design task is rarely a blank sheet; it is rather a parametric variation of proven geometries under changed requirements, such as a larger jaw width, a different steel grade, or revised handle ergonomics. A design agent instantiates such a variant from the product model, generates candidate geometries, and evaluates them against strength, mass, and manufacturability; deep generative models for this synthesis-and-evaluation loop are established in engineering design research. Simulation is invoked as a tool rather than performed as a human activity: the agent parameterizes the finite-element model of a forging step or a torque load

case, submits it, reads the result and iterates. A downstream agent checks the candidate against the available forging dies, machining centers, and heat-treatment recipes and returns the process plan together with the geometry.

Manufacturing is the part of the Dark Company for which the closest empirical analog exists. Forming and machining of the reference product family are already executed by numerically controlled equipment; forging presses, transfer lines, grinding and polishing cells as well as continuous heat-treatment furnaces run on programmed cycles, and optical or tactile stations classify the results. What the agent layer adds is not the actuation but the decision: order sequencing and line balancing against due dates and tool availability, the release of process-parameter corrections when inspection data drift, and the re-planning of affected orders when a cell fails.

Intralogistics and material handling closes the gap between these cells. Robotic handling at the machine, autonomous transport between the forging, machining, heat-treatment, and finishing areas, and an automated small-parts warehouse together make the material flow addressable by software: every movement becomes a request an agent can issue and confirm. The demanding element here is not transport but the bulk handling of small, geometrically similar, and often oiled parts, where gripping and singulation remain the practical bottleneck and where the physical rather than the cognitive layer determines feasibility.

Maintenance is the department in which the concept is most exposed, because it is where the physical world degrades. Condition monitoring and the prediction of impending failures by machine learning belong to the most thoroughly reviewed Industry 4.0 applications and may be considered available in principle; an agent that schedules an intervention before a spindle bearing or a die fails is a planning problem of the kind agents handle well. The step the concept requires beyond this is the execution of intervention without human hands: robots that change dies and tool inserts, exchange filters and wear parts and, in the limit, repair the machines that produce. This is the least mature element of the configuration and the point at which a fully dark hand-tool plant currently fails.

Purchasing is a candidate for near-complete delegation. The demand side follows from the order backlog and the production plan; the supply side, for a company that buys steel in defined grades, coatings, and standard components, is a bounded set of qualified suppliers. A procurement agent forecasts demand, release orders against those contracts, requests and compares quotes, and conducts bounded negotiations within an explicit mandate of price and lead-time limits, qualified-supplier lists, and escalation thresholds. Supplier qualification and the award of a new frame contract are the decisions that do not follow from the data, because they concern counterparty risk and relationships rather than optimization, and they are therefore retained above the automation threshold.

Sales is where the revenue claim of the concept must be settled. For a catalog business, quote and order intake are largely a mapping problem: an inquiry arriving by e-mail, electronic data interchange, or web form is parsed, matched to catalog items and configurable variants, priced against the customer's contract conditions and volume tiers, checked for availability against the production plan and answered. This is the workflow of Fig. 2 applied to a commercial rather than a service transaction, and order confirmation, delivery scheduling, invoicing, and dunning follow the same pattern. Non-catalog requests, custom tooling, and framework negotiations with key accounts exceed the pattern and are routed to the escalation path.

Marketing becomes a generation and scheduling function. Campaign concepts, copy and imagery, product content for distributor portals and catalogs, and the continuous operation of social-media channels can be produced and published by agents at a cadence that no small marketing department sustains. The retained controls are brand consistency and legal review, covering competition law, product claims, and image rights, and they are enforced as guardrails on publication rather than as manual authorship.

The support functions divide along the same line. Quality management is largely data work: inspection records, capability indices, complaint handling, and the documentation required by the applicable product standards can be compiled and evaluated by agents, whereas accreditation and the signature on a declaration of conformity remain attributable to a person. Accounting and finance follow from the transaction data that the agents themselves generate, which makes continuous rather than periodic reporting the natural mode. Human resources becomes residual by construction, since in the limiting configuration there are few employees to administer. Information technology and compliance, by contrast, grow in importance and change in character, from supporting human users to operating the agent fleet and managing its identities, permissions, model versions, cost limits, and audit trails.

What remains for management is orchestration and the assumption of responsibility. Objectives, budgets, mandate boundaries, and escalation thresholds are set by humans, and the agent layer executes within them. Coordination across agents is itself a governed function, comprising a supervisor pattern that decomposes and assigns tasks, an explicit protocol for inter-agent communication, a persistent record of what was decided on which basis, and a guardrail layer that intercepts high-risk or ambiguous cases. Multi-agent research supplies the mechanisms for this and documents at the same time its characteristic failure mode, the propagation of an early error through a chain of nominally independent agents. In the limiting configuration considered here, a single founder specifies the system once and thereafter verifies periodically that it still runs, which is a supervisory rather than an operational role and is precisely what decouples output from headcount.

4 Discussion and Conclusions

The core economic assumption - that agent operation opens up previously unprofitable, low-margin or fine-grained business fields - holds in its basic logic but requires differentiation. Substituting personnel with computational power opens the "long niche"; at the same time, inference is not free, and falling barriers to entry let competition erode the very margins that were unlocked. The bottleneck thus migrates from headcount to orchestration, error control, and defensibility. Further key hurdles are hallucinations and cascading errors, the poisoning of scraping agents through "fake data injection", the ESG footprint (environmental, social, and corporate governance), and the gap between demonstrator and stable operation.

The feasibility of the configuration is uneven across departments, and the unevenness is systematic. Where the object of work is data and the action is a transaction, as in quotes, order processing, release order, reporting, and content production, the required capabilities exist today, and the remaining gap is one of integration and governance. Where the object of work is matter, feasibility is bounded by the physical layer: machining, forming, heat treatment, and inspection of a stable hand-tool geometry are solved; the bulk handling of small, oiled, and geometrically similar parts is not reliably solved. The concept is therefore not uniformly speculative: its office half is an integration problem, its shop-floor half is largely an available capability, and its maintenance half is an open research problem. Assessments that treat the Dark Company as a single feasibility question connect the three.

In regulatory terms, the EU AI Act imposes tight limits: with broad applicability from August 2026, fines of up to € 35 million or 7% of worldwide annual revenue are possible, and the duties arising from the deployment of such a system attach to the operating undertaking and its management rather than to the system itself; final responsibility remains with the human operator. A complete decoupling of the enterprise from humans therefore remains, in the European context, an operational and legal illusion. The viable configuration is partial autonomy, in which human experts act as strategic orchestrators and the final control instance, flanked by robust governance, and agent-based identity management. Of particular relevance is therefore the extent to which such structures open up previously unprofitable business fields, and raise efficiency without incurring incalculable legal and operational risks.

Liability follows from this asymmetry rather than from the degree of automation. An agent is not a legal person, and no jurisdiction relevant to the reference company grants it that status; the obligations attaching to a product placed on the market remain with the undertaking and with its management. Accountability frameworks make the requirement explicit, in that attributability, answerability, and auditability all presuppose an identifiable principal who can be called upon to account, which is a property of the organization and not of the agent. This is what bounds the limiting case. A single founder who verifies periodically that the system still runs is legally the operator of everything it does, which turns the residual human role from a convenience into a structural requirement: someone must remain able to know, and to answer for, what the system decided.

Three limits therefore delimit the concept. Technically, autonomous physical maintenance is absent, and without it a plant is dark only until something wears out. Organizationally, the volume of exceptions decides the outcome: the reference company was chosen because its products are stable and its orders repetitive, whereas a business with unstable specifications or negotiated one-off contracts would generate an exception rate that consumes the saved effort in escalation. Legally and economically, the decoupling is real but partial, since output no longer scales with

headcount yet still scales with compute, with integration effort, and with the capacity of a few people to remain accountable for a system they no longer operate. The Dark Company is therefore best read not as a forecast but as a limiting construct: a way of asking which functions of a manufacturing enterprise are bound to human labor by necessity rather than by convention. For the hand-tool manufacturer considered here the answer is a small set, comprising supplier and customer relationships that carry counterparty risk, decisions that create legal obligations, and the physical repair of the means of production.

Acknowledgement

AI tools were used to assist with language refinement, condensing text, formatting, and generating figures. All content is based on the authors' own research, writing, and manually compiled sources, and was fully reviewed and verified editorially by the authors.

References

[1] Wang, R., Tong, Y., Zhuang, C. (2025): Lights-out factories: Review and prospect. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 239(14), p. 1983-1993, doi: 10.1177/09544054241305826

[2] Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., Wen, J. (2024): A survey on large language model based autonomous agents. Frontiers of Computer Science 18(6), id. 186345, doi: 10.1007/s11704-024-40231-1

[3] Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., Wiest, O., Zhang, X. (2024): Large Language Model Based Multi-agents: A Survey of Progress and Challenges. Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24), p. 8048-8057, doi: 10.24963/ijcai.2024/890

[4] van der Aalst, W. M. P., Bichler, M., Heinzl, A. (2018): Robotic Process Automation. Business & Information Systems Engineering 60(4), p. 269-272, doi: 10.1007/s12599-018-0542-4

[5] Xu, X., Lu, Y., Vogel-Heuser, B., Wang, L. (2021): Industry 4.0 and Industry 5.0—Inception, conception and perception. Journal of Manufacturing Systems 61, p. 530-535, doi: 10.1016/j.jmsy.2021.10.006

[6] Leitão, P. (2009): Agent-based distributed manufacturing control: A state-of-the-art survey. Engineering Applications of Artificial Intelligence 22(7), p. 979-991, doi: 10.1016/j.engappai.2008.09.005

[7] Regenwetter, L., Nobari, A. H., Ahmed, F. (2022): Deep Generative Models in Engineering Design: A Review. Journal of Mechanical Design 144(7), id. 071704, doi: 10.1115/1.4053859

[8] Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., Li, G. P. (2020): Predictive maintenance in the Industry 4.0: A systematic literature review. Computers & Industrial Engineering 150, id. 106889, doi: 10.1016/j.cie.2020.106889

[9] Acemoglu, D., Restrepo, P. (2020): Robots and Jobs: Evidence from US Labor Markets. Journal of Political Economy 128(6), p. 2188-2244, doi: 10.1086/705716

[10] Novelli, C., Taddeo, M., Floridi, L. (2024): Accountability in artificial intelligence: what it is and how it works. AI & Society 39(4), p. 1871-1882, doi: 10.1007/s00146-023-01635-y

OpenUSD as a Semantic Framework for GMP-Oriented and Validatable Automated Optical Inspection

Radmer, Jochen[1]
[1]Jade Hochschule, Friedrich-Paffrath-Str. 101, 26389 Wilhelmshaven, jochen.radmer@jade-hs.de

Abstract: Automated optical inspection (AOI) enables 100 % visual quality control. Regulatory requirements (*Good Manufacturing Practice* (GMP) and data integrity (DI)) in pharmaceutical manufacturing also demand complete traceability of image and process data throughout the entire product lifecycle. The heterogeneity of the vast amount of data poses the greatest challenge here. In this article, we investigate OpenUSD as a hierarchical scene graph format for use as a unified data and semantic layer to consistently integrate heterogeneous AOI data and support regulatory traceability within a validated overall system. A reference architecture design, a compact mapping table, and an exemplary code snippet demonstrate the feasibility of a conceptual approach. Practical GMP validation and performance evaluation will be the subject of future work.

Keywords: Open Universal Scene Description, OpenUSD , automatisierte optische Inspektion, GMP, DI, Semantik

1 Introduction

Automated optical inspection is indispensable in pharmaceutical production because it objectifies visual quality control. To this end, AOI systems capture geometric and visual characteristics throughout the production process and must ensure validated procedures, controlled changes, and robust documentation in order to simultaneously meet *Good Manufacturing Practice* (GMP) and data integrity (DI) requirements throughout the entire lifecycle [1–5]. For the implementation of GMP, *EU GMP Annex 11* also requires validation, audit trails, access controls, secure data storage, and data recovery for computerized systems [2]. To ensure data integrity according to ALCOA+, records must be traceable, legible, timely, authentic, accurate, complete, consistent, durable, and available [4, 6]. These requirements are considered system requirements in this paper. OpenUSD provides a data representation for this purpose but does not replace a validated operating environment, access control, or archiving solution. The data considered in this context includes image data, process parameters, machine status, release decisions, deviations, and validation protocols, which are generated in various formats, systems, and life cycle phases [1–5, 7]. The greatest challenge lies in this heterogeneity of data. This raises the question of how these data can be combined to meet the diverse requirements. There is therefore a need to consolidate this data consistently so that every inspection decision remains traceable throughout the entire equipment and product lifecycle, thereby creating a robust audit trail.

For AOI systems, this means that image and metadata streams, calibration parameters, and model/recipe versions must be archived in a traceable and immutable manner. In addition, AOI pipelines, data models, and configurations must undergo computer system validation (CSV) (CSV)—from the User Requirements Specification (URS) through risk-based assessment to Design Qualification (DQ), Installation Qualification (IQ), Operational Qualification (OQ), and Performance Qualification (PQ)—and be able to document changes in a traceable manner [3, 8].

To this end, this paper aims to present an approach to consistent representation and integration that can bridge these gaps. To this end, the following sections outline the fundamentals of AOI in the pharmaceutical context and highlight the current state of the art and the specific challenges in regulated production lines. This paper makes three key contributions. First, it presents an analysis of the regulatory requirements for AOI in pharmaceutical manufacturing. Second, a OpenUSD -based reference architecture is designed that integrates not only semantics, geometry, and temporal relationships but also the qualification artifacts (URS, DQ, IQ, OQ, PQ), thereby creating a consistent, audit-ready database. Third, a concise mapping table is provided that maps each GMP requirement to a possible OpenUSD mechanism and the required supplementary controls.

2 Current Status and Challenges

This fragmentation is reinforced by inconsistent calibration data and coordinate frames, as proprietary formats, varying units, and inconsistent identifiers make it difficult to link CAD references, inspection features, and measurement data [9]. In regulated pharmaceutical environments, however, calibrations, measurement uncertainties, and time references must be documented in a controlled manner and maintained in a traceable manner via audit trails so that inspection results remain reliable for batch decisions [2, 7]. This requires consistent 3D references across packaging materials, the system, the facility, and camera systems—including lighting—as well as the maintenance of provenance metadata, authors, and timestamps in accordance with ALCOA+ to capture quality and trace the causes of quality changes [4].

In practice, AOI systems are distributed across separate subsystems and are only loosely coupled via industrial middleware, which leads to semantic disconnects between design data, inspection logic, and production status. Against this background, specific challenges can be identified:

- Heterogeneous data sources: CAD models, image streams, process parameters, and AOI parameterizations/models are available in different formats, units, time frames, and lifecycles and lack consistent identifiers [4].
- Variant diversity and batch reference: Product families, packaging materials, and special versions require flexible configurations, robust tolerance and attribute management across variant spaces, and a unique link to lot, batch, and serialization.
- Versioning and change control: Every change (e.g., new calibration data) must be traceable as a standalone, audit-ready commit.
- Integration in a regulated real-time environment: AOI pipelines require production-synchronous latencies and a complete audit trail
- Validation of machine learning-based methods. Models, datasets, and inference pipelines require controlled changes, traceable approvals, and scheduled revalidations to ensure that results remain reproducible and suitable for batch decisions [2, 8].
- Security and role management: Only authorized persons may change configurations. Changes must be cryptographically signed.

These challenges call for a consistent representation and integration layer that brings together semantics, geometry, temporal relationships, and metadata, thereby enhancing digital connectivity.

3 OpenUSD und alternative Datenmodelle

OpenUSD defines a *Stage* model in which entities of the digital cell—such as cameras, components, and test features—can be modeled as hierarchically organized *Prims*. Changes are stored in *Layers*, and multiple layers can be combined into a *Layer Stack*. *TimeSamples* enable the storage of time-dependent attribute values at defined *TimeCode* coordinates. The mapping to real-world timestamps, time scales, and synchronization sources must be specified separately. OpenUSD is already integrated into CAD and rendering tools (Autodesk, Siemens NX, NVIDIA Omniverse) [10].

Relevant alternative data models include:

- OPC-UA Information Model (IEC 62541): OPC UA focuses on communication and information models for industrial systems, while OpenUSD provides a hierarchical scene graph with composition, layer, and geometry mechanisms.
- ROS-URDF / SDF: XML-based formats for describing robot kinematics, sensors, and collision geometry. Focused on articulated systems, but without an integrated audit trail or OpenUSD-like layer composition mechanisms, and not intended for generic industrial assets.

- MTConnect: an open XML standard for capturing manufacturing data (ISO/TC 184). It provides time-ordered data and simple hierarchical asset identification; it is not primarily designed for a hierarchical 3D scene graph or the combined composition of heterogeneous engineering and inspection artifacts.

Unlike alternative data models, OpenUSD offers a comprehensive, versionable, and semantically rich data model, as it combines a hierarchical scene graph, layer-based change management and versioning in conjunction with an external version control system (VCS), time samples, extensible schemas, and efficient payload processing, as well as interoperability with engineering tools [10]. Therefore, we are investigating OpenUSD as a potential representation layer within a GMP-validated overall system.

4 Implementation of Regulatory GMP/DI Requirements based on OpenUSD

Relevant requirements for AOI systems can be derived from the regulatory guidance documents and are listed in Tables 1 [1–5]. The tables also outline technical approaches for implementing the requirements based on OpenUSD. This mapping should be understood as an architectural hypothesis. An actual audit may require additional services, procedures, and evidence. OpenUSD serves as the representation and integration layer and does not replace validated GMP processes or external control systems.

Table 1: GMP-/DI-Anforderungen und OpenUSD-Unterstützung für AOI-Systeme

#	**GMP/DI Requirements**	**OpenUSD Support and Supplementary Checks**
A1	URS	Requirements as a user-defined, versioned `aoi:` schema with typed attributes and relationships. Formal management and approval via a requirements management system.
A2	Func. Spec.	Functional elements as hierarchical `UsdPrim`s. Interfaces and dependencies via `UsdRelationship`s. Pipeline steps as user-defined Prims. This does not implement an execution logic.
A3	Design Spec.	System architecture as a stage hierarchy with typed prims for hardware, sensors, coordinate systems, and configurations. Design approval via a controlled document system.
A4	Risk	Risks, measures, and affected components via user-defined attributes and `UsdRelationship`s. Evaluation and approval within a validated risk management process.
A5	Qualification	Reference URS, DQ, IQ, OQ, and PQ artifacts via stable IDs and relationships. Test results and deviations as external, addressable artifacts. Formal qualification outside of OpenUSD.
A6	Change Control	Structure changes using layers, layer stacks, and edit targets. `UsdEditContext` supports the selection of the edit target. Audit trail and approvals via a validated external system.
A7	DI	Typed attributes, stable IDs, provenance information, and external object references for data and digests. Data integrity, immutability, and WORM protection through validated archiving and storage systems.
A8	Audit Trail	Reference affected objects, layer/commit IDs, and external audit events via attributes and relationships. `UsdTimeCode` can represent time references but is not a complete audit timestamp. Implement an audit trail externally with old/new values, user, timestamp, and reason.
A9	Access	Link user and role references to an external IAM/RBAC system using persistent IDs. OpenUSD metadata documents these mappings.
A10	Backup	Export layers and stages using `UsdStage::Export()` and transfer them to a backup system. Does not implement a validated backup and restore procedure itself. Backup, restore, retention, and recovery testing are performed by a validated external system.
A11	Security	Document security status and policy references as user-defined attributes. Implement access control, key management, network management, and vulnerability management outside of OpenUSD.
A12	Signatures	Reference signed external artifacts via signature ID, certificate reference, and digest. Signature and key management handled externally.

Table 2: GMP-/DI-Anforderungen und OpenUSD-Unterstützung für AOI-Systeme

#	GMP/DI Requirements	OpenUSD Support and Supplementary Checks
A13	System Doc	Reference documents via stable IDs, versions, and relationships to Prims. Generate reports using external analysis services
A14	Periodic Review	Model review objects, time periods, statuses, and reviewed versions as Prims, attributes, and relationships. Planning, execution, and approval through a validated review process. Formal review approval takes place outside of OpenUSD.
A15	Env. Controls	Model environmental data as user-defined measurement Prims with unit, device ID, calibration reference, and time reference. Acquisition, alerting, and storage via validated measurement and process systems.
A16	Calibration	Model calibrations as user-defined prims with parameters, units, uncertainties, validity periods, and certificate references. Versioning via layers and IDs. External process monitoring.
A17	CSV	Reference lifecycle status, configuration versions, and validation artifacts via attributes and relationships. Support traceability; test execution, deviations, release, and revalidation through validated CSV processes.

5 Reference architecture

At the heart of the reference architecture is a central OpenUSD stage that digitally models the entire test environment, including the production cell, the product under test, and all connected sensors. These elements are organized hierarchically within a modular and scalable structure that can be easily adapted to different production facilities. The data structure is implemented using layers, with each layer having a specific lifecycle and a defined function (see Fig. 1).

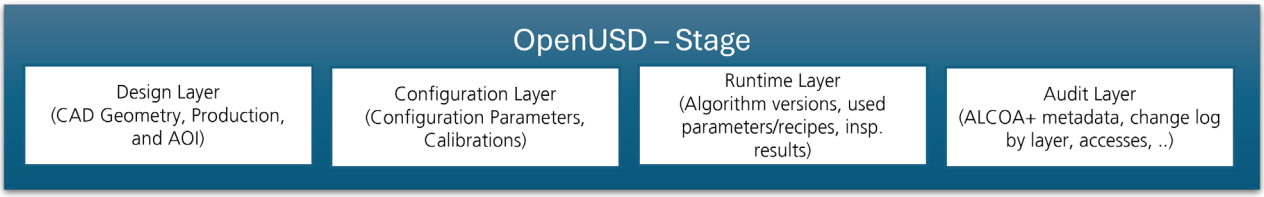


Fig. 1: Reference architecture: OpenUSD stage with layers for Design, Configuration, Run Time, and Audit, accessed by the AOI pipelines for writing (RT) and reading (RT, Conf., and Design).

The *Design Layer* contains the original, unmodified CAD models of the machines, components, and sensors. This foundation is static and serves as an unchanging basis for the entire digital cell as long as it remains unchanged. This requires a review of the tolerances of products and packaging materials. The *Configuration Layer* contains configurable data that does not change the geometry but influences behavior. This includes, for example, camera settings, inspection tolerances, or variable sensor poses. Versioning allows relationships between settings and specific points in time to be traced. The *Run Time Layer* contains dynamic data generated during the running process. Here, time-related inspection events, measured values, and references to image data are stored as typed attributes or *Prims*. The actual image data can be archived in an external object store and referenced via stable IDs or content IDs. *Time-Samples* allow time-dependent attribute values to be structured and trends to be analyzed. The mapping to actual process times must be defined separately. The *audit layer* contains referencing metadata on changes and links the affected OpenUSD objects to external audit events. The authoritative audit trail is recorded in an external, validated system. The layer may also contain immutable copies of relevant change logs or their digests. The documented information includes, for example, the creator (`di:createdBy`), the creation date (`di:createdAt`), the details of the change, and a checksum value (`di:checksum`).

The AOI pipeline uses the central OpenUSD stage and accesses various layers. The *Design Layer* imports CAD geometries with PLM metadata as an immutable foundation, while the *Configuration Layer* maintains versioned calibration matrices and tolerance limits specific to product variants to meet change control requirements. During the runtime phase, time-related inspection events, measured values, and references to image data are stored as OpenUSD attributes or prims. The actual image data can be archived in a suitable external object store and referenced via stable IDs or content IDs. The *audit layer* references the changes in a versioned layer stack. The immutability and authenticity of the authoritative records are ensured by external, validated archive and signature controls. The AOI pipeline accesses this geometry via *payload* references and can thus not only document the AOI system but also, in the future, detect discrepancies when comparing real and synthetic data through image and lighting simulations (e.g., for synthetic data).

The OpenUSD code in Fig. 2 shows how OpenUSD can be used as a semantic representation layer within a GMP-validated AOI system. The combination of hierarchical paths (for traceability), user-defined attributes (for semantics and production data), and DI metadata results in a consistent data model that can support regulatory audit trail requirements within a correspondingly validated overall system. In terms of its structure, the approach can be adapted to multiple inspection characteristics and batches. The actual scalability and runtime performance must be investigated experimentally. It should be noted that OpenUSD itself does not provide complete audit trail, IAM, signature, archiving, or CSV functionality. These functions must be implemented by external, validated systems.

An example CSV traceability chain links user requirement `URS-REQ-001` to risk analysis `RISK-004`, design qualification `DQ-012`, installation qualification `IQ-TEST-006`, the operational qualification `OQ-TEST-017`, the performance qualification `PQ-TEST-023`, the test result `RESULT-017`, and the release `RELEASE-003`. OpenUSD stores persistent IDs and references for these; the authoritative qualification and release documents remain in validated source systems.

```
from pxr import Usd, UsdGeom, Sdf

stage = Usd.Stage.CreateInMemory()
feat = UsdGeom.Xform.Define(stage, "/Part/Features/Hole01").GetPrim()

feat.CreateAttribute("aoi:featureType", Sdf.ValueTypeNames.Token).Set("cyl_hole")
feat.CreateAttribute("aoi:tolValue_mm", Sdf.ValueTypeNames.Double).Set(0.05)
feat.CreateAttribute("aoi:batchId", Sdf.ValueTypeNames.String).Set("BATCH-2020-047")
feat.CreateAttribute("di:createdBy", Sdf.ValueTypeNames.String).Set("Max_Mustermann")
feat.CreateAttribute("di:createdAt", Sdf.ValueTypeNames.String).Set("2020-12-10T12:15:00Z")
```

Fig. 2: OpenUSD-Code: Examples of test characteristics (namespace `aoi:`) with data integrity metadata (`di:`)

The OpenUSD code in Fig. 3 illustrates integration into existing production networks—*Manufacturing Execution System* (MES) and *Product Lifecycle Management* (PLM)—via *OPC-UA*. Communication with actuators and sensors can take place via *ROS-2*. To this end, the proposed data model uses *adapter prims* that contain unique identifiers (e.g., `ua:nodeId`, `ros:topic`). The attributes document the external binding, and the actual communication is implemented by dedicated OPC-UA and ROS-2 adapters. Thanks to *lazy loading* of payloads, only the subgraphs relevant to the current production run are loaded into memory.

```
from pxr import Sdf

# part is a previously defined UsdPrim
# Binding to OPC-UA
part.CreateAttribute("ua:nodeId", Sdf.ValueTypeNames.String).Set("ns=2;s=Cell/PartA/State")
# Binding to ROS-2
part.CreateAttribute("ros:topic", Sdf.ValueTypeNames.String).Set("/cell/part_a/state")
```

Fig. 3: OpenUSD Code: Integration with production networks via OPC-UA (namespace ua) and ROS-2 (namespace ros)

6 Conclusion and Outlook

This article presented an architectural design for a uniform and versionable representation layer that can support traceability for GMP- and DI-relevant data within a validated overall system. The conceptual reference architecture and the compact mapping schema presented here enable AOI manufacturers to get started quickly. The conceptual analysis demonstrates the potential of this approach. Its regulatory suitability and performance—particularly with regard to scaling and modularization concepts such as *lazy loading*—must be demonstrated through practical implementation, risk analysis, and validation. These are the next steps.

This work presents a plausible architectural concept. However, a statement regarding regulatory compliance requires implementation, risk analysis, and validation—all of which are still pending. It enables a seamless connection between production practices, data analysis, and compliance.

References

[1] World Health Organization, *Quality assurance of pharmaceuticals: a compendium of guidelines and related materials. Volume 1: Good practices and related regulatory guidance*. Geneva: World Health Organization, 10th ed., 2024.

[2] European Commission, "Eudralex volume 4 eu guidelines for good manufacturing practice annex 11 computerised systems," tech. rep., 2011. Revision 1.

[3] European Commission, "Eudralex volume 4 eu guidelines for good manufacturing practice annex 15 qualification and validation," tech. rep., 2015.

[4] Food and Drug Administration, "Data integrity and compliance with drug cgmp: Questions and answers guidance for industry," guidance for industry, U.S. Department of Health and Human Services, Dec. 2018.

[5] World Health Organisation, "Trs 1033 - annex 4: Who guideline on data integrity," tech. rep.

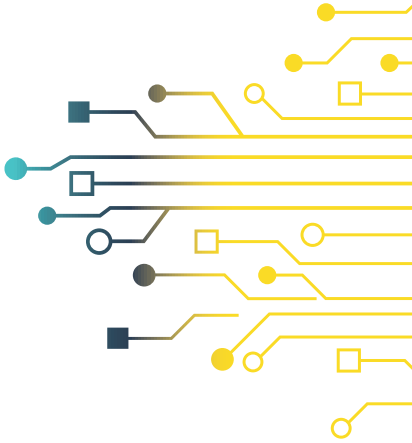
[6] World Health Organization, "Who technical report series no. 996 annex 5 guidance on good data and record management practices," tech. rep., 2016.

[7] PIC/S Pharmaceutical Inspection Cooperation Scheme, "Pi 041-1 good practices for data management and integrity in regulated gmp/gdp environments," tech. rep., 2021.

[8] International Society for Pharmaceutical Engineering (ISPE), *GAMP 5 A Risk-Based Approach to Compliant GxP Computerized Systems Second Edition*. ISPE, 2022.

[9] S. Catalucci, A. Thompson, S. Piano, D. Branson, and R. Leach, "Optical metrology for digital manufacturing: a review," *The International Journal of Advanced Manufacturing Technology*, vol. 119, pp. 1–34, 2022.

[10] Pixar Animation Studios, "Universal scene description (usd) – specification," 2021.



Über uns:

Die Gesellschaft zur Förderung angewandter Informatik e. V. (GFaI) ist eine industrienahe Forschungseinrichtung mit über 30 Jahren Erfahrung in der angewandten Informatik. Wir entwickeln praxisorientierte IT-Lösungen für Unternehmen und Institutionen, insbesondere in den Bereichen optische Mess- und Prüftechnik, Schall und Schwingungen, 3D-Datenverarbeitung, Robotik, Energie- und Umwelttechnik, Netzwerkinfrastruktur sowie wissensbasierte Dialogsysteme.

Neben der Forschung engagiert sich die GFaI in der Ausbildung und Qualifizierung von Fachkräften. In Zusammenarbeit mit Hochschulen fördern wir den Wissenstransfer und begleiten Studierende sowie Nachwuchstalente auf ihrem Weg in die Praxis.

Zusammen mit ihren Tochterunternehmen gfai tech GmbH und Adalogic GmbH unterstützt die GFaI den Transfer von Innovationen in den Mittelstand Deutschlands und setzt Impulse für digitale und nachhaltige Entwicklungen.

Weitere Informationen finden Sie unter www.gfai.de.

Kontaktdaten

GFaI Future Tech Days
E-Mail: futuretechdays@gfai.de
Internet: www.gfai.de/aktuelles/veranstaltungen/gfai-future-tech-days/workshop

IMPRESSUM

Herausgeber / Organisation:
F. Böhm, B. Hohnhäuser, 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: sekretariat@gfai.de
Internet: www.gfai.de

ISBN-Nummer:
978-3-942709-38-5