

USING ADVANCED IMAGE ANALYSIS AND 3D MODELLING TO ENSURE THE QUALITY AND SAFETY OF GENE THERAPEUTICS IN AN ISOLATED ENVIRONMENT

**Nikola Cudráková, Petr Strakoš, Petra Svobodová,
Milan Jaroš, Jaroslav Krutí, Ondřej Zbránek,
Petr Lesný**

Abstract

Maintaining consistent quality in pharmaceutical manufacturing is especially challenging in the manual production of gene therapeutics, where operator interactions within isolator environments play a crucial role. Traditional methods rely on manual documentation or basic video monitoring using single or multi-camera setups. Such approaches are limited in accuracy and scalability. This study proposes a novel framework that replaces conventional image analysis with advanced computer vision and real-time 3D modeling. By reconstructing operator actions in three dimensions, the system offers enhanced spatial precision and a richer interpretation of quality-relevant parameters. This capability enables more accurate monitoring of critical tasks, supports traceability, and strengthens process integrity. Through selected use cases, we demonstrate how this approach can improve the reliability of operator tracking and introduce new possibilities for quality assurance and auditing in the production of gene therapies. The results highlight the transformative potential of 3D computer vision technologies in regulated pharmaceutical environments.

Keywords

gene therapeutics, isolated environment, advanced image analysis, 3D modeling

1 Introduction

Manufacturing of gene therapeutics is typically performed under Good Manufacturing Practice (GMP) controls in cleanrooms, restricted-access barrier systems (RABS) or isolators, where personnel interactions are tightly constrained to minimize contamination risks. The manufacturing process of gene therapeutics in isolators needs to utilize complex devices, such as an electroporator, a centrifuge, or automatic pipettes. These devices require careful observation for their function and proper setup. The latest revision of EU GMP Annex 1 emphasizes a holistic Contamination Control Strategy (CCS), upgraded environmental monitoring, and the use of appropriate technologies such as isolators and robotics for sterile manufacturing principles that directly influence how operator tasks are performed and verified in practice [1]. In parallel, FDA's Aseptic Processing guidance explicitly addresses personnel qualification, cleanroom design, environmental monitoring, and the use of isolators, underlining the critical role of human interventions in sterile operations [2,3]. Gene therapy production adds further complexity: small batch sizes, manual manipulations, and variable workflows make consistency difficult and magnify the impact of operator technique on quality and safety [4,5].

Traditional documentation (paper/electronic batch records) and basic video monitoring offer limited spatial fidelity and are often retrospective, constraining timely detection of deviations and weakening traceability of complex manual tasks. Regulatory expectations for trustworthy, audit-ready electronic records (21 CFR Part 11) increase the burden on manufacturers to ensure data integrity, attribution, and complete audit trails across operator-driven steps [6,7]. Moreover, contamination control

scholarship and industry practice repeatedly identify personnel interventions and airflow disturbances around operators as dominant risks. Airflow visualization and computational fluid dynamics analyses show how human motion can perturb unidirectional flow and particle transport in critical zones. These effects are hard to evaluate from single-view 2D footage [8,9]. In cell and gene therapy settings, these gaps are exacerbated by frequent product changeovers, bespoke methods, and scale-out rather than scale-up, all of which strain conventional monitoring approaches [4].

Advanced computer vision (CV) has transformed quality control in other pharmaceutical operations (e.g., packaging and visual inspection), demonstrating robust defect detection and in-line analytics [10]. However, most deployments focus on product surfaces rather than a precise 3D understanding of human actions in aseptic workspaces. Recent progress in multi-view, real-time 3D human pose estimation (3D HPE) enables accurate, low-latency reconstruction of human kinematics from multiple cameras, even under occlusions, and opens a pathway to quantify operator posture, hand trajectories, and tool-material interactions in three dimensions [11–13]. When fused with process metadata, 3D reconstructions could map movements to GMP-relevant actions, verify adherence to standard operating procedures (SOP), and contextualize events against airflow or environmental monitoring data to explain deviations better. Aligning such CV-derived records with quality requirements of FDA CFR Part 11 and EudraLex strengthens data integrity, traceability, and review activities [6,7].

This work introduces a framework that can replace conventional, essentially 2D image analysis with multi-camera advanced CV and real-time 3D modeling of operator actions within an isolator. Specifically, we aim to: (i) reconstruct operator kinematics in 3D with sufficient spatial precision for GMP-relevant tasks; (ii) extract quality-relevant parameters (e.g., hand-vial distances, exposure durations, transfer paths) aligned to SOP steps; (iii) integrate outputs with CCS elements to better understand contamination risk; and (iv) generate reviewable, Part 11-supportive records that enhance traceability and auditability. By grounding the approach in current regulatory guidance [1,2,6] and leveraging state-of-the-art real-time 3D HPE [11–13], we target measurable improvements in monitoring accuracy and process integrity for the manual production of gene therapies.

2 Related Work

Current industrial practice increasingly integrates automated visual inspection (AVI) into aseptic and isolator lines as part of pharmaceutical process monitoring, combining multi-camera imaging, controlled lighting, and AI/ML classifiers to detect visible particles, container/closure defects, cosmetic flaws, and fill-volume errors. These systems support 100% inspection and can replace or complement manual inspection when properly qualified. Regulators explicitly encourage the use of “appropriate technologies” and continuous monitoring systems in a sterile production context that has accelerated AVI adoption and its tighter linkage with batch release decisions [14].

In traditional aseptic processing, many operations are still carried out by hand inside cleanrooms or isolators. Hand pose estimation and tracking, together with general objects detection and tracking, are extensively researched areas in CV. Hand pose estimation in 2D or 3D is typically performed using RGB [15], RGB-D, or depth images as input. A similar approach holds for reliable object detection [16]. Depth information is invaluable because it provides extra details about the shape and distance of objects. This makes pose estimation and object de-

tection more reliable, especially when body parts and objects are occluded or when the background is cluttered.

In recent years, deep learning has greatly advanced hand pose estimation. A typical pipeline includes three steps [17, 18]. First, a hand detection algorithm (e.g., YOLO [19], MediaPipe palm detector [20], or transformer-based detector [21]) is used to localize the hand region within an image or video frame. The result is usually a bounding box or segmentation mask. In the second step, 2D keypoints are extracted to mark the anatomical landmarks of the hand. Frameworks such as OpenPose and MediaPipe are widely used for this task, as they can generate reliable skeleton representations that serve as the basis for subsequent action recognition. Finally, 3D hand pose estimation converts the 2D landmarks into a complete 3D skeleton of the hand. This works best with depth or RGB-D input, which directly provides the missing third dimension.

Pose estimation is also widely used as input for action recognition or hand-object interaction detection [22]. Pose features capture the motion and articulation of the body or hands, while RGB frames provide additional context, such as the objects being held or the surrounding scene [23]. Depth information, when available through RGB-D sensors, adds valuable geometric cues that make recognition more robust under occlusion and viewpoint changes. Combining pose, RGB, and depth has been shown in many studies [24, 25] to improve action recognition performance significantly.

Across high-precision industries, 3D modelling is already central to quality assurance. In semiconductors, manufacturers use volumetric non-destructive testing [26] (e.g., X-ray microscopy, scanning acoustic microscopy, terahertz microscopy) to provide precise detection and quantification of defects in manufactured components, thereby enabling more reliable, automated quality assurance and process optimization. In aerospace, high-resolution X-ray CT is routine for 3D exploration and modeling to detect internal defects in carbon fibre reinforced polymer (CFRP) [27]. At the same time, digital-twin methods can fuse 3D assemblies with live shop-floor data to assess geometric deviation and predict quality during product assembly [28]. Emerging human digital twins (HDTs) extend this idea by reconstructing operator actions from multi-camera video and VR pipelines to provide continuous, objective conformity feedback [29]. For biotech and sterile manufacturing, these examples show how multi-view 3D reconstructions (of parts or people) can quantify tolerances, exposure paths, and intervention timing—principles that translate directly to isolator-based CGT workflows to strengthen traceability and deviation analysis.

3 Materials and Methods

3.1 Prototype of the isolator

The prototype of the isolator is designed as a single chamber with two gloves, see Figure 1, with the possibility of vaporized hydrogen peroxide (VHP) sterilization. The basic dimensions of the chamber are 850×1002×850 mm. The material of the inner chamber is made of AISI 316L with a ground surface to reduce glare. The chamber is equipped with two powerful LED lights, which are located in the ceiling of the cabin and increase the overall visibility inside the chamber. Inside the cabin is a custom-developed gripping system that can mount various types of cameras. The mentioned system can position the cameras at any position along the side walls and ceiling in 6 DoF.

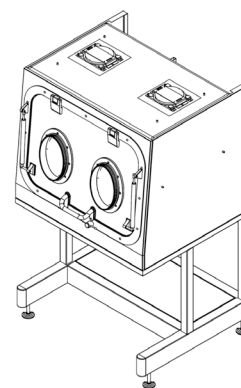


Figure 1 – Prototype of the isolator with a single chamber and two gloves

3.2 3D model of the isolator

A prototype of the isolator has been designed in 3D software and, as such, offers the use of the isolator 3D model in a virtual environment. Besides that, the Department of Immunotherapy of the Institute of Hematology and Blood Transfusion, which routinely manufactures products for clinical trials using aseptic technology inside an isolator, has provided detailed 3D models of items inside the isolator. 3D models of equipment, such as Petri dishes, centrifuge tubes, expansion flasks, or electroporation stations, were placed accordingly into the virtual isolator with respect to the SOP for the manufacture of T cells with chimeric antigen receptors (CAR-T).

3.3 3D Reconstruction pipeline

To monitor the activities performed in the isolator, the detection and 3D tracking of objects present in the chamber have to be handled. Each laboratory equipment and raw material that enters the isolator needs to be uniquely labeled in order to be recognized and its presence in the isolator approved. To ensure that, 2D codes, such as QR, Data Matrix, etc., can be used. For that purpose, we develop and implement approaches that can identify objects based on their shape and unique 2D code labeling. YOLO object detectors are used together with 2D code readers provided by the pylibdmtx [30] and pyzbar [31] libraries.

Special attention has to be given to recognition of hand actions in glovebox environments, where operators perform tasks while wearing gloves and interacting with laboratory equipment. This setting introduces additional challenges, including frequent occlusions and the fact that the hands may be partially covered by tools or equipment inside the glovebox. To address these challenges, we capture synchronized RGB and depth video using cameras placed in the glovebox. The RGB stream provides visual and contextual cues, while the depth sensor provides geometric information that enables precise hand localization and 3D pose estimation. Our method consists of three main stages.

1. Hand Detection/Segmentation – In the first step, hands in gloves are localized within each video frame using detection or segmentation models such as YOLO, MediaPipe palm detector, or transformer-based detectors. The result is a bounding box or mask that isolates the hands from the rest of the scene. Detecting gloved hands is a challenging task, since gloves lack the natural texture and structure of skin, and existing models are usually trained on bare-hand datasets. To address this, we plan to collect data relevant to our glovebox environment and fine-tune existing models to improve robustness.

2. 3D Pose Estimation – Using the depth channel, we can estimate the 3D pose of the detected hands. This describes their position and movement in space, which is necessary for analyzing actions inside the glovebox. Depth information makes the estimation more reliable, especially when the hands are covered by gloves, holding tools, or partially hidden by the glovebox structure. The resulting 3D representation provides information that can be directly used in the action recognition stage.

3. Action/Interaction Recognition – Having 3D poses available, we can analyze how they change over time to recognize the actions performed. Activities will be classified by analyzing the sequence of 3D hand poses together with contextual RGB-D information, such as the presence of tubes or other tools in the glovebox. This combination allows us to capture both the motion of the hands and their interaction with objects over time. The reconstructed 3D hand and tool trajectories can be mapped to predefined motion templates aligned with GMP-relevant tasks, such as aseptic transfers, pipetting, or vial handling. From these trajectories, the system can extract quality parameters, for example, movement precision, exposure times, and distances to critical biological materials, and automatically flag deviations from SOP-defined tolerances, enabling objective detection of operator errors or contamination risks.

To further improve robustness in cluttered glovebox environments, we will employ a multi-stereo camera setup, with, e.g., one stereo camera capturing the hands from the right and another from the left. Combining these viewpoints will make it possible to reduce occlusions, since a hand that is partially covered by a tool in one view may still be visible in the other. This setup will also support more precise 3D pose estimation, as depth information can be obtained from multiple perspectives and fused into a more reliable representation.

3.4 Data acquisition and preprocessing

In the proposed framework, data acquisition and preprocessing can leverage modern stereo vision platforms such as Luxonis cameras with DepthAI API or Stereolabs ZED cameras with ZED SDK, both of which enable real-time 3D reconstruction in constrained environments. These systems support multi-view synchronization and calibration, allowing fusion of data streams to mitigate occlusions caused by operator posture or equipment. Advanced preprocessing, including reflection filtering under isolator lighting and adaptive background subtraction, is essential to maintain robust tracking. Particular attention must be given to personal protective equipment, mainly the gloves, which obscure natural skin and joint features; hand-pose models can be retrained or fine-tuned on gloved hand datasets, combined with depth cues, to ensure reliable detection of critical manipulations despite uniform textures or glare.

4 Results

4.1 Reading of 2D codes

Objects in the isolator are marked with 2D codes (e.g., Data Matrix and QR codes), which must be detected and decoded. Data Matrix codes were read using the Python library `pylibdmtx` [30], while QR codes were decoded using `pyzbar` [31]. The Dynamsoft Barcode Reader SDK [32] was also tested as an alternative commercial solution. All tested libraries produced comparable results.

A key factor for successful code detection and decoding is camera configuration, which results in pixel density per object size at a specific distance. Experiments showed that each module of a 2D code needs a minimum number of pixels to be reli-

ably read. For example, if we have a 1×1 cm Data Matrix code with 14×14 modules, at least 2 pixels per module are needed, and depending on the camera field of view (FOV), resolution, and distance of the code from the camera, these are either met or not. Increasing the number of pixels per module improves detection reliability but reduces the distance from the camera. The experimental results, including the maximum reading distances for each code type, library, and camera configuration, are summarized in Tables 1 and 2.

Experiments compared two cameras with different resolutions and image processing. Using a Luxonis camera (OAK-D PRO W PoE) [33] with resolution 1280×800 , 1×1 cm Data Matrix codes (14×14 modules) could be reliably read at distance up to 13 cm, whereas a higher-resolution camera with narrower FOV, the Obsbot meet 2 with 1920×1080 resolution [34] achieved larger reading distance (up to 65 cm) with the same 1×1 Data Matrix code.

Type of code	1.0x1.0 cm	1.3x1.3 cm	2.0x2.0 cm
Data matrix (pylibdmtx)	13 cm	11 cm	23 cm
QR code (pyzbar)	7 cm	13 cm	25 cm
Data matrix (DynamSoft)	13 cm	19 cm	30 cm
QR code (DynamSoft)	9 cm	13 cm	26 cm

Table 1 – Maximum reading distances for Data Matrix and QR codes using the Luxonis camera with 1280×800 resolution

Type of code	1.0x1.0 cm	1.3x1.3 cm	2.0x2.0 cm
Data matrix (pylibdmtx)	65 cm	70 cm	110 cm
QR code (pyzbar)	45 cm	70 cm	90 cm
Data matrix (DynamSoft)	60 cm	70 cm	110 cm
QR code (DynamSoft)	35 cm	60 cm	80 cm

Table 2 – Maximum reading distances for Data Matrix and QR codes using the Obsbot Meet 2 camera with 1920×1080 resolution

4.2 Identification and 3D tracking of objects

We use a one-stage YOLO detector to identify laboratory tools in each video frame. The detector is trained on photos of the target objects from the laboratory using the standard augmentations to capture appearance variability. At runtime, it operates in real time and returns the class, bounding box, and confidence score.

After the identification of the laboratory objects, the ZED SDK takes the left and right RGB images, rectifies them, estimates disparity, and converts them into a depth map. From each detection, we extract a reliable depth estimate and reconstruct its 3D coordinates (X, Y, Z) and then track these coordinates over time.

The baseline (distance between the two cameras) directly controls depth accuracy and range – a larger baseline yields lower depth error at the same distance. Camera FOV also matters. A narrower FOV acts like a longer lens, so objects occupy more pixels and depth estimates become more precise, see Figure 2. With multi-view fusion of calibrated, time-synchronized ZED cameras, the depth error is typically lower than the single-rig values as

Recommended zone (error < 5%)

Depth not available

GS NARROW

Distance		10cm	30cm	50cm	0.75cm	1m	1.5m	2m	3m	4m	5m	7m	10m	20m	25m	50m	100m	200m	300m	
Baseline	3cm		1mm	2mm	5mm	9mm	20mm	3cm	8cm	14cm	22cm	43cm	87cm							
	7cm		<1mm	1mm	2mm	4mm	8mm	1cm	3cm	6cm	9cm	18cm	37cm	1.49m	2.33m					
	15cm				1mm	2mm	4mm	1cm	2cm	3cm	4cm	9cm	17cm	69cm	1.09m	4.35m				
	25cm					1mm	2mm	<1cm	1cm	2cm	3cm	5cm	10cm	42cm	65cm	2.61m	10.45m			
	50cm							<1cm	<1cm	1cm	1cm	3cm	5cm	21cm	33cm	1.30m	5.21m	20.90m		
	75cm								<1cm	1cm	1cm	2cm	3cm	14cm	22cm	87cm	3.47m	13.91m	31.35m	
	1m									<1cm	1cm	1cm	3cm	10cm	16cm	65cm	2.61m	10.43m	23.48m	
	1.5m										1cm	2cm	7cm	11cm	43cm	1.74m	6.95m	15.64m		
	2m												1cm	5cm	8cm	33cm	1.30m	5.21m	11.73m	

Figure 2 – Stereolabs baseline-distance tab. Estimated depth accuracy at various distances for a single stereo system with narrow FOV lenses [35].

provided in Figure 2. The improvement comes from combining independent views, and that reduces random noise [35].

4.3 Scene reconstruction from 3D model and captured data

Using the obtained models, we have assembled a virtual scene where we can perform experiments similar to those performed in the real isolator environment. We can either use captured data from the cameras, such as 3D trajectories of the objects with relevance to each other, and revisit them in the virtual environment, or simulate operations in the virtual scene and capture the events using virtual cameras that can then be used as inputs to detection algorithms (object detection/tracking, 2D codes reading, hand pose estimation).

In Figure 3, we provide an example from the created virtual isolator environment. The scene contains a complete 3D model of the isolator, including proper material assignment, together with selected 3D models of items from the isolator environment.

In Figure 4, we utilize the 3D virtual isolator environment and generate video feeds from the virtual cameras to obtain RGB

and depth maps of the visible scene. This information is used as input to the detection algorithms to test their performance. The whole system can work interactively, meaning the camera view in the scene can be changed based on the user's preference, and the detection algorithms are applied immediately afterwards and provide near real-time feedback to the user.

5 Discussion

The ability of dynamically recreating a 3D model of the environment inside an isolator, where a human operator is carrying out the aseptic manufacturing operation, allows us to increase the accuracy of intervention detection, enhance contamination risk assessment, and provide objective, reviewable records that strengthen overall GMP compliance and process understanding. When compared to classical 2D image analysis, which is limited to pixel-level or frame-based evaluation of events, a semantically enriched 3D scene model provides a much deeper understanding of the environment. By combining the depth information, object recognition, tracking of operator movements, and a knowledge database containing descriptions of manufacturing processes, such models can contextualize operator actions in real time.

This semantic layer transforms raw video data into structured knowledge that can be used for automated risk scoring, predictive contamination control, and even personnel training by replaying and annotating interventions in a virtualized environment. Moreover, because the 3D reconstruction captures the spatial relationships between operators, equipment, and sterile products, it supports proactive decision-making. It facilitates regulatory review by offering objective, high-level representations of aseptic practices.

Moreover, this setup can dramatically reduce the bandwidth required for remote isolator monitoring (only the recognized actions and 3D spatial information of the operator's hands are transmitted).

In this sense, semantic modeling represents the next evolutionary step beyond conventional video monitoring, bridging the gap between passive recording and intelligent, GMP-compliant process understanding.

The software created in this way will need to be valid under 21 CFR Part 11, i.e., meet the requirements for electronic records (usually generated as a text log-file) and signatures to ensure their reliability, security, and validity in accordance with the rules of the US FDA. Key features are: audit trails that record all changes, secure electronic signatures that are unique to the user and legally binding, user access control to ensure authorization, data encryption, and system validation that proves that the software reliably functions according to its intended purpose.

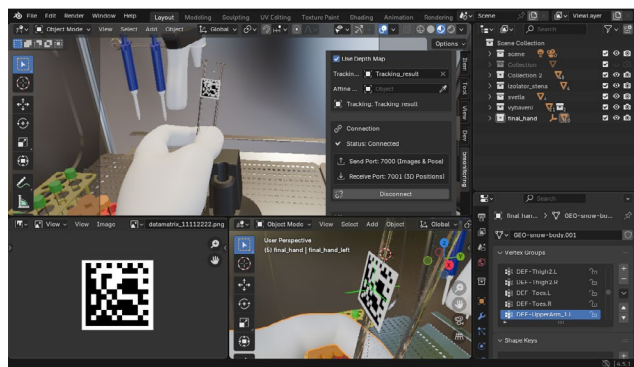


Figure 3 – Virtual isolator environment created in Blender software



Figure 4 – Using a virtually generated camera feed in detection algorithms (object detection/tracking, 2D code reading)

6 Conclusion

This study demonstrates the feasibility and benefits of using advanced computer vision and 3D modeling to monitor operator actions in isolator-based gene therapy manufacturing. By combining multi-view imaging, depth data, and semantic scene reconstruction, the framework will enable precise tracking of laboratory tools, materials, and operator hand movements, even under personal protective environment-induced variability and occlusion. Compared to conventional 2D monitoring, the 3D models will provide a richer, contextual understanding of aseptic practices, supporting real-time deviation detection, contamination risk assessment, and traceable audit records. This capability will not only enhance GMP compliance and process integrity but also create new opportunities for predictive quality assurance, remote monitoring, and operator training.

Acknowledgements

This work was co-financed from the state budget by the Technology Agency of the Czech Republic and the Ministry of Industry and Trade of the Czech Republic under the TREND Programme.

References

- [4.] European Commission. EU GMP Annex 1: Manufacture of Sterile Medicinal Products (2022, Rev. 1). https://health.ec.europa.eu/system/files/2022-08/20220825_gmp-an1_en_0.pdf.
- [5.] U.S. FDA. Sterile Drug Products Produced by Aseptic Processing – Current Good Manufacturing Practice. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/sterile-drug-products-produced-aseptic-processing-current-good-manufacturing-practice>.
- [6.] U.S. FDA. Guidance for Industry: Sterile Drug Products Produced by Aseptic Processing. <https://www.fda.gov/media/71026/download>.
- [7.] Fury, Brian, and Gerhard Bauer. "GMP manufacturing of cell and gene therapy products: Challenges, opportunities, and pathways forward." *Molecular Therapy* (2025).
- [8.] Bozenhardt H.F., Bozenhardt E.H. Selecting The Right Isolator For Cell Therapy Manufacturing: Considerations & Challenges (2020). <https://www.cellandgene.com/doc/selecting-the-right-isolator-for-cell-therapy-manufacturing-considerations-challenges-0001>.
- [9.] 21 CFR Part 11 – Electronic Records; Electronic Signatures (eCFR). <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-11>.
- [10.] U.S. FDA. Part 11, Electronic Records; Electronic Signatures – Scope and Application. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/part-11-electronic-records-electronic-signatures-scope-and-application>.
- [11.] Air visualization study – A Comprehensive Overview (2025). <https://www.golighthouse.com/en/wp-content/uploads/2025/04/Air-Visualization-Studies.pdf>.
- [12.] Guldana, Abiyeva, et al. "Impact of Airflow Disturbance from Human Motion on Contaminant Control in Cleanroom Environments: A CFD-Based Analysis." *Buildings* 15.13 (2025): 2264.
- [13.] Galata, Dorian Laszlo, et al. "Applications of machine vision in pharmaceutical technology: A review." *European Journal of Pharmaceutical Sciences* 159 (2021): 105717.
- [14.] Ye, Hang, et al. "Faster voxelpose: Real-time 3d human pose estimation by orthographic projection." *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022.
- [15.] Ci, Hai, et al. "Proactive multi-camera collaboration for 3d human pose estimation." *arXiv preprint arXiv:2303.03767* (2023).
- [16.] Boldo, Michele, et al. "Real-time multi-camera 3D human pose estimation at the edge for industrial applications." *Expert Systems with Applications* 252 (2024): 124089.
- [17.] Scheme, Pharmaceutical Inspection Co-operation. "Guide to Good Manufacturing Practice for Medicinal Products.", Rev. (2023).
- [18.] C. Zimmermann and T. Brox. Learning to estimate 3D hand pose from single RGB images. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [19.] Chen, Ai, et al. "Advancing in RGB-D salient object detection: A survey." *Applied Sciences* 14.17 (2024): 8078.
- [20.] C. Zimmermann and T. Brox. Learning to estimate 3d hand pose from single rgb images. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [21.] P. Panteleris, I. Oikonomidis, and A. A. Argyros, "Using a single RGB frame for real time 3D hand pose estimation in the wild," *arXiv preprint arXiv:1712.03866*, 2017.
- [22.] J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger. In *Computer Vision and Pattern Recognition, 2017. CVPR 2017. IEEE Conference on*, 2017.
- [23.] Sánchez-Brizuela, G., Císal, A., de la Fuente-López, E. et al. Lightweight real-time hand segmentation leveraging MediaPipe landmark detection. *Virtual Reality* 27, 3125–3132 (2023). <https://doi.org/10.1007/s10055-023-00858-0>.
- [24.] Cho, Hoseong, et al. "Transformer-based unified recognition of two hands manipulating objects." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023.
- [25.] M. Schröder and H. Ritter. Hand-object interaction detection with fully convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017.

- [26.] D. C. Luvizon, D. Picard and H. Tabia, "Multi-Task Deep Learning for Real-Time 3D Human Pose Estimation and Action Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2752–2764, 1 Aug. 2021.
- [27.] Hu, Jian-Fang, et al. "Deep bilinear learning for rgb-d action recognition." *Proceedings of the European conference on computer vision (ECCV)*. 2018.
- [28.] Khaire, Pushpajit, Javed Imran, and Praveen Kumar. "Human activity recognition by fusion of rgb, depth, and skeletal data." *Proceedings of 2nd International Conference on Computer Vision & Image Processing: CVIP 2017, Volume 1*. Singapore: Springer Singapore, 2018.
- [29.] Su, Yutai, et al. "Volumetric nondestructive metrology for 3D semiconductor packaging: A review." *Measurement* 225 (2024): 114065.
- [30.] Dilonardo, Elena, et al. "High resolution X-ray computed tomography: A versatile non-destructive tool to characterize CFRP-based aircraft composite elements." *Composites Science and Technology* 192 (2020): 108093.
- [31.] Zhuang, Cunbo, et al. "Digital twin-based quality management method for the assembly process of aerospace products with the grey-markov model and apriori algorithm." *Chinese Journal of Mechanical Engineering* 35.1 (2022): 105.
- [32.] Colaw, Christopher Lee, et al. "Methodology for Enablement of Human Digital Twins for Quality Assurance in the Aerospace Manufacturing Domain." *Sensors* 25.11 (2025): 3362.
- [33.] Natural History Museum. pylibdmtx. GitHub, (2025), <https://github.com/NaturalHistoryMuseum/pylibdmtx/tree/master>.
- [34.] Natural History Museum. pyzbar. GitHub, (2025), <https://github.com/NaturalHistoryMuseum/pyzbar>.
- [35.] Dynamsoft. Barcode Reader Overview. Dynamsoft, (2025), <https://www.dynamsoft.com/barcode-reader/overview/>.
- [36.] Luxonis. OAK-D Pro W/PoE. (2025), <https://shop.luxonis.com/products/oak-d-pro-w-poe>.
- [37.] OBSBOT. OBSBOT Meet 2 4K Webcam. (2025), <https://www.obsbot.com/obsbot-meet-2-4k-webcam?srltid=AfmBOoryvc7PrvFs-FRYT3-AG3Yhpse6MfjLZIRwHfXUqloQQyZJcbpcE>.
- [38.] Stereolabs. Setting up your Stereo Rig with two ZED X One cameras. Stereolabs Support, (2025), <https://support.stereolabs.com/hc/en-us/articles/20851232732311-Setting-up-your-Stereo-Rig-with-two-ZED-X-One-cameras>.

Kontakt

Petr Strakoš

IT4Innovations
národní superpočítačové centrum
VŠB – Technická univerzita Ostrava
Studentská 6231/1B
708 00 Ostrava-Poruba
petr.strakos@vsb.cz

Nikola Cudráková

IT4Innovations
národní superpočítačové centrum
VŠB – Technická univerzita Ostrava
Studentská 6231/1B
708 00 Ostrava-Poruba
nikola.cudrakova@vsb.cz