Augmented Reality in the Operating Room for Neurosurgical Interventions

Christian Kunz[1], Franziska Mathis-Ullrich[1], and Björn Hein[1,2]

[1] Karlsruhe Institute of Technology (KIT), Institute for Anthropomatics and Robotics, Karlsruhe, Germany
franziska.ullrich@kit.edu
[2] Karlsruhe University of Applied Sciences, Karlsruhe, Germany
bjoern.hein@hs-karlsruhe.de

Abstract. Neurosurgical procedures are associated with great challenges for the surgeon since a high degree of precision is required. Operations are performed within a limited space and often concealed structures are not visible to the surgeon. A system is proposed that integrates augmented reality into a digital operating room. The basis for this is an understanding of the scene and the integration into the surgical workflow. In a first step a two-stage process is implemented to detect the patient on the operating table with high precision. Further a solution is presented to semantically segment the surgical scene to detect and track medical instruments. For better understanding of the situation in the operation room the medical staff is tracked with OpenPose. These solutions build the base for a precise and robust integration of augmented reality into the digital operating room.

Keywords: Computer Assisted Surgery, Augmented Reality, Neurosurgical Interventions, Digital OR, Ventricular Puncture.

1 Introduction

Computer-assisted technologies in medical interventions are intended to support the surgeon during treatment and improve the outcome for the patient. One possibility is to augment reality with additional information that would otherwise not be perceptible to the surgeon. In medical applications, it is particularly important that demanding spatial and temporal conditions are adhered to. Challenges in augmenting the operating room are the correct placement of holograms in the real world, and thus, the precise registration of multiple coordinate frames to each other, the exact scaling of holograms, and the performance capacity of processing and rendering systems.

In general, two different scenarios can be distinguished. First, applications exist, in which a placement of holograms with an accuracy of 1 cm and above are sufficient. These are mainly applications where a person needs a three-dimensional view of data. An example in the medical field may be the visualization of patient data, e.g. to understand and analyse the anatomy of a patient, for diagnosis or surgical planning. The correct visualization of these data can be of great benefit to the surgeon. Often only 2D patient data is available, such as CT or MRI scans. The availability of 3D representations depend strongly on the field of application. In neurosurgery 3D views are available but often not extensively utilized due to their limited informative value. Additionally computer monitors are a big limitation, because the data can not be visualized in real world scale. Further application areas are the translation of known user interfaces into augmented

Fig. 1. A neurosurgeon performing a ventricular puncture wearing a HoloLens.

reality (AR) space. The benefit here is that a surgeon refrains from touching anything, but can interact with the interface in space using hand or voice gestures. Applications visualizing patient data, such as CT scans, only require a rough positioning of the image or holograms in the operation room (OR). Thus, the surgeon can conveniently place the application freely in space. The main requirement is then to keep the holograms in a constant position. Therefore, the internal tracking of the AR device is sufficient to hold the holograms at a fixed position in space. The second scenario covers all applications, in which an exact registration of holograms to the real world is required, in particular with a precision below 1 cm. These scenarios are more demanding, especially when holograms must be placed precisely over real patient anatomy. To achieve this, patient tracking is essential to determine position and to follow patient movements. The system therefore needs to track the patient and adjust the visualization to the current situation. Furthermore, it is necessary to track and augment surgical instruments and other objects in the operating room. The augmentation needs to be visualized at the correct spatial position and time constraints need to be fulfilled. Therefore, the AR system needs to be embedded into the surgical workflow and react to it. To achieve these goals modern state of the art machine learning algorithms are required. However, the computing power on available AR devices is often not yet sufficient for sophisticated machine learning algorithms. One way to overcome this shortcoming is the integration of the AR system into a distributed system with higher capabilities, such as the digital operating theatre OP:Sense (see Fig. 2).

In this work an augmented reality system *HoloMed* [4] (see Fig. 1) is integrated into the surgical research platform for robot assisted surgery *OP:Sense* [5]. The objective is to enable high-quality and patient-safe neurosurgical procedures in order to increase the surgical outcome by providing surgeons with an assistance system that supports them in cognitively demanding operations. The physician's perception limits are extended by

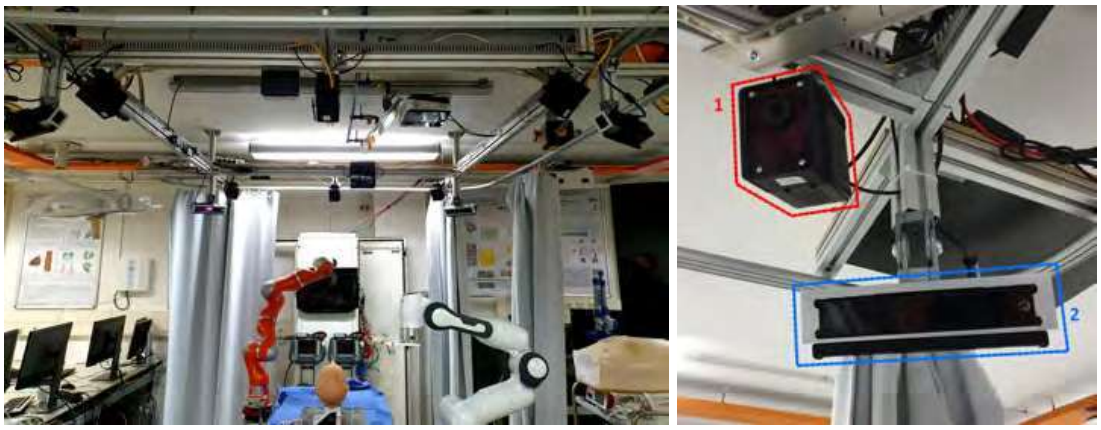
the AR system, which bases on supporting intelligent machine learning algorithms. AR glasses allow the neurosurgeon to perceive the internal structures of the patient's brain. The complete system is demonstrated by applying this methodology to the ventricular puncture of the human brain, one of the most frequently performed procedures in neurosurgery. The ventricle system has an elongated shape with a width of 1-2 cm and is located in a depth of 4 cm inside the human head. Patient models are generated fast ($<$ 2s) from CT-data [3], which are superimposed over the patient during operation and serve as a navigation aid for the surgeon. In this work the expanded system architecture is presented to overcome some limitations of the original system where all information were processed on the Microsoft HoloLens, which lead to performance deficits. To overcome these shortcomings the HoloMed project was integrated into OP:Sense for additional sensing and computing power.

2 Material and Methods

To achieve integration of AR into the operation room and the surgical workflows, the patient, the instruments and the medical staff need to be tracked. To track the patient, a marker system is fixated on the patient head and registration from the marker system to the patient is determined. A two-stage process was implemented for this purpose. First the rough position of the patient's head is determined on the OR table by applying a YOLO v3 net to reduce the search space. Then a robot with a mounted RGB-D sensor is used to scan the acquired area and build a point cloud of the same. To determine the patient's head in space as precisely as possible a two-step surface matching approach is utilized. During recording, the markers are also tracked. With known position of the patient and the markers, the registration matrix can be calculated. For the ventricular puncture a solution is proposed to track the puncture catheter to determine the depth of insertion into the human brain. By tracking the medical staff the system is able to react to the current situation, e.g. if an instrument is passed. In the following the solutions are described in detail.

2.1 Experimental Setup

Our digital operating room OP:Sense (illustrated in Fig. 2a) consists of an OR table with two robots attached to it, a Kuka LWR4 and a Franka Panda lightweight robot. Several sensors are integrated into the setup on the ceiling rack: an ARTRRACK 2 system to track retroreflective markers consisting of six IR cameras and four Microsoft Kinect sensors (shown in Fig. 2b). Any objects can be tagged with markers to track them, provided that marker-to-instrument registration is available. Robot and operating table can be tracked in the room via ART markers. The Microsoft Kinect sensors provide a RGB-D stream of the operation area. Intel RealSense D415 and D435 cameras can additionally be placed inside the OR or can be mounted on the robots, to capture a defined near-field area. A patient phantom head from Synbone and a custom-build phantom skull were used during experiments. A Microsoft HoloLens is used to visualize AR to the surgeon. It employs the Unity 3D graphics engine to visualize the holographic scene. Patient tracking is provided through two different marker systems: 1) the Aruco library in combination with OpenCV and 2) the Vuforia library. OP:Sense is based on the Robot Operating System (ROS), which is a middleware for robotic platforms, consisting of a set of software libraries and tools. Core components are so-called nodes connecting all system components with each other.



(a) (b)

Fig. 2. a) OP:Sense system setup. b) Sensor setup with six ARTTRACK 2 IR cameras (red 1) and four Microsoft Kinect sensors (blue 2).

2.2 Patient Detection

To detect the patient's head, the coarse position is first determined with the YOLO v3 CNN [6], performed on the Kinect RGB image streams. The position in 3D is determined through the depth stream of the sensors. The OR table and the robots are tracked with retroreflective markers by the ARTTRACK system. This step reduces the spatial search area for fine adjustment. The Franka Panda has an attached Intel RealSense RGB-D camera as depicted in Fig. 3.

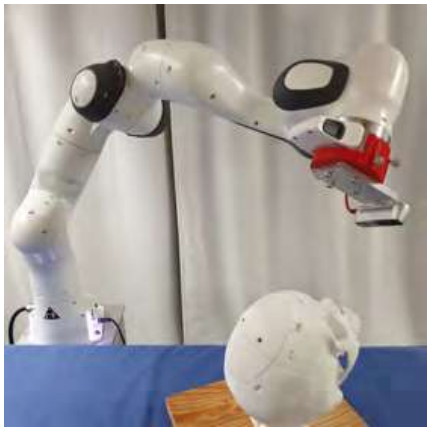


Fig. 3. Franka Panda with attached RGB-D sensor.

The precise determination of the position is performed on the depth data with surface matching. The robot scans the area of the coarsely determined position of the patient's head. A combined surface matching approach with feature-based and ICP matching was implemented. The process to perform the surface matching is depicted in Fig. 4. In clinical reality, a CT scan of the patient head is always performed prior to a ventricular puncture for diagnosis, such that we can safely assume the availability of CT data. A process to segment the patient models from CT data was proposed by Kunz et al. in [3]. The algorithm processes the CT data extremely fast in under two seconds. The data format is '.nrrd', a volume model format, which can easily be converted into surface models or

point clouds. The point cloud of the patient's head CT scan is the reference model that needs to be found in OR space. The second point cloud is recorded from the RealSense depth stream mounted on the Panda robot by scanning the previously determined rough position of the patient head. All points are recorded in world coordinate space. The

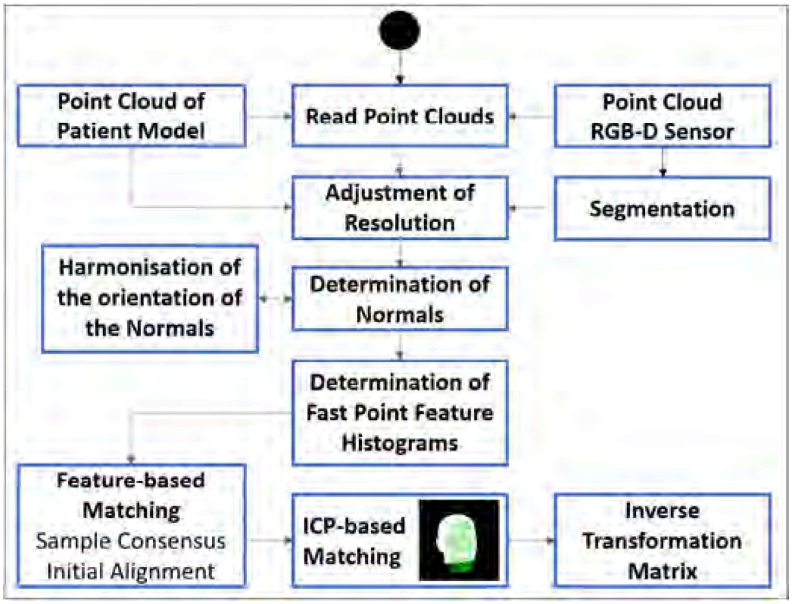


Fig. 4. Surface matching process.

search space is further restricted with a segmentation step by filtering out points that are located on the OR table. Additionally, manual changes can be made by the surgeon. In a performance optimization, the resolution of the point clouds is reduced to decrease processing time without loosing too much accuracy. The normals of both point clouds generated from CT data and from the recorded RealSense depth stream are subsequently calculated and harmonised. During this step, the harmonisation is especially important as the normals are often misaligned. This misalignment occurs because the CT data is a combination of several individual scans. For alignment of all normals, a point inside the patient's head is chosen manually as a reference point, followed by orienting all normals in the direction of this point and subsequently inverting all normals to the outside of the head (see Fig. 5).

After the preprocessing steps, the first surface fitting step is executed. It is based on the Initial alignment algorithm proposed by Rusu et al. [8]. An implementation within the point cloud library (PCL) is used. Therefore fast point feature histograms need to be calculated as a preprocessing step. In the last step an iterative closest point (ICP) algorithm is used to refine the surface matching result. After the two point clouds have been aligned to each other the inverse transformation matrix can be calculated to get the correct transformation from marker system to patient model coordinate space.

2.3 Catheter Tracking

As outlined in Fig. 6, catheter tracking was implemented based on semantic segmentation using a Full-Resolution Residual Network (FRRN) [7]. After the semantic segmentation of the RGB stream of the Kinect cameras, the image is fused with the depth stream

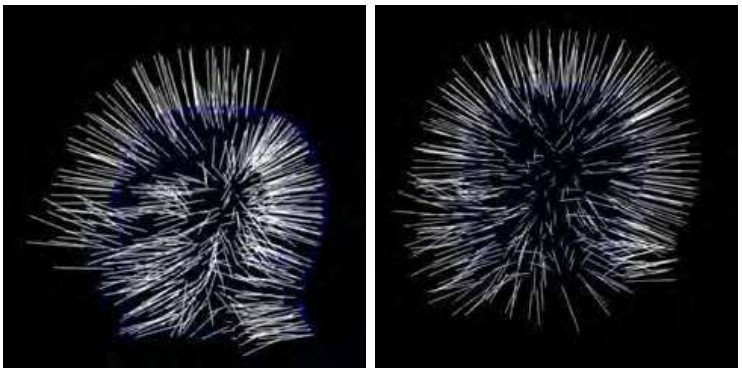


Fig. 5. Before (left) and after (right) harmonisation of wrongly oriented normals.

to determine the voxels in the point cloud belonging to the catheter. As a further step a density based clustering approach [2] is performed on the chosen voxels. This is due to noise especially on the edges of the instrument voxels in the point cloud. Based on the found clusters an estimation of the three dimensional structure of the catheter is performed. For this purpose, a narrow cylinder with variable length is constructed. The length is changed accordingly to the semantic segmentation and the clustered voxels of the point cloud. The approach is applicable to identify a variety of instruments.

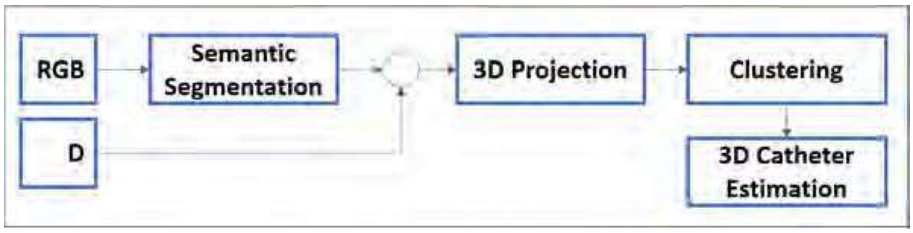


Fig. 6. Process to track the catheter using a RGB-D sensor.

2.4 Tracking of Medical Staff

The OpenPose [1] library is used to track key points on the bodies of the medical staff. Available ROS nodes have been modified to integrate OpenPose in the OP:Sense ROS environment. The architecture is outlined in Fig. 7.

3 Results

In this chapter the results of the patient, catheter and medical staff tracking are described. The approach to find the coarse position of a patient's head was performed on a phantom head placed on the OR table within OP:Sense. Multiple scenarios with changing illumination and occlusion conditions were recorded. The results are depicted in Fig. 8 and the evaluation results are depicted in Table 1.

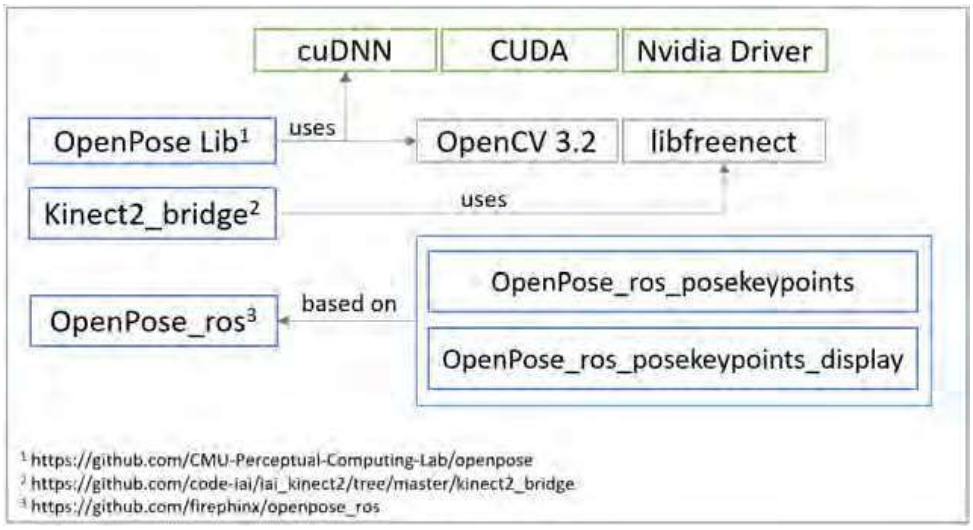


Fig. 7. Integration of the OpenPose library into the ROS environment of OP:Sense.

Fig. 8. Coarse determination of the patient's head on the OR table with YOLO v3.

Table 1. Evaluation results for the phantom search with YOLO v3.

	Precision	Recall	F1-Score	average IoU	mAP
Normal OR conditions	92%	99%	95%	67.59%	90.35%
Occlusion	99%	93%	96%	75.77%	90.86%
Strong Illumination	62%	66%	64%	41.01%	62.23%
Illumination and Occlusion	65%	51%	57%	41.83%	45.62%

Precision detection of the patient was performed with a two-stage surface matching approach. Different point cloud resolutions were tested with regard to runtime behaviour. Voxel grid edge sizes of 6, 4 and 3 mm have been tested, with a higher edge size corresponding to a smaller point cloud. The matching results of the two point clouds were analyzed manually. An average accuracy of 4.7 mm was found with an accuracy range between 3.0 and 7.0 mm.

In the first stage of the surface matching, the two point clouds are coarsely aligned as depicted in Fig. 9. In the second step ICP is used for fine adjustment. A two-stage process was implemented as ICP requires a good initial alignment of the two point clouds.

Table 2. Evaluation results of the two-step surface matching.

Voxel Grid Size (mm)	Points	Points after Adjustment of Resolution	Processing Time (in minutes)	Mean Accuracy (mm)
6	17418	1293	1.98	6.45
4	17418	2706	7.57	3.52
3	17418	4405	17.87	4.12
4	47530	2860	7.53	3.94
4	49917	2774	7.71	3.0
4	30521	3035	8,09	7.0

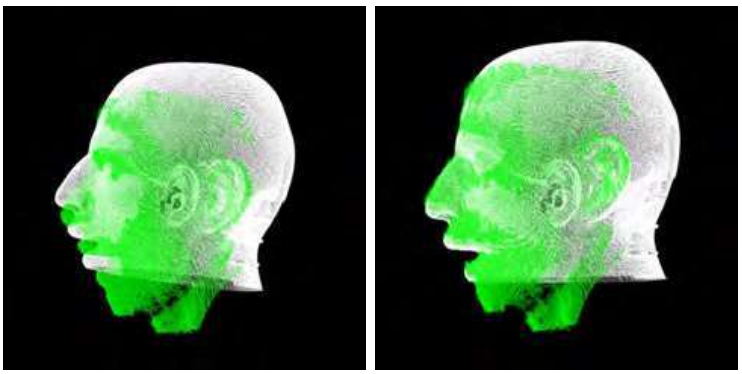


Fig. 9. Results of the two-stage surface matching. Left: After Initial Alignment, Right: After ICP.

For catheter tracking a precision of the semantic segmentation between 47% and 84% is reached (see Table 3). Tracking of instruments, especially neurosurgical catheters, are challenging due to their thin structure and non-rigid shape. Detailed results on catheter tracking have been presented in [7].

Table 3. Evaluation results for semantic segmentation [7].

Dataset	Normal conditions 1	Normal conditions 2	Catheter horizontal	Catheter vertical	Catheter diagonal	Bright	Bright, patient covered
Precision	84.1%	64.0%	77.1%	47.7%	81.0%	59.0%	72.9%
Recall	58.7%	61.9%	51.2%	19.2%	31.0%	15.4%	43.1%

The 3D estimation of the catheter is shown in Fig. 10. The catheter was moved in front of the camera and the 3D reconstruction was recorded simultaneously. Over a long period of the recording over 90% of the catheter are tracked correctly. In some situations this drops to under 50% or lower.

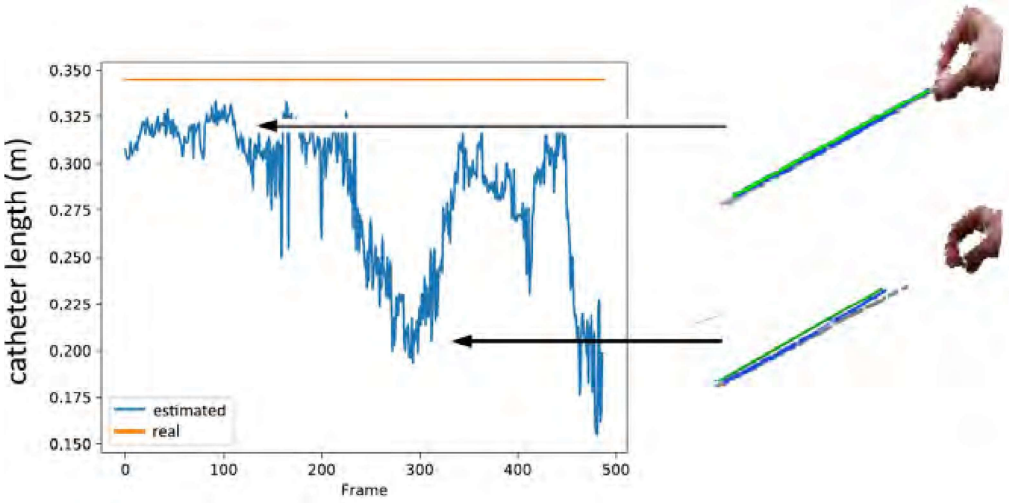


Fig. 10. Results of the 3D estimation of the catheter.

The tracking of medical personnel is shown in Fig. 11. The different body parts and joint positions are determined, e.g. the head, eyes, shoulders, elbows, etc. The library yielded very good results as described in [1]. We reached a performance of 21 frames per second on a workstation (Intel i7-9700k, GeForce 1080 Ti) processing 1 stream.

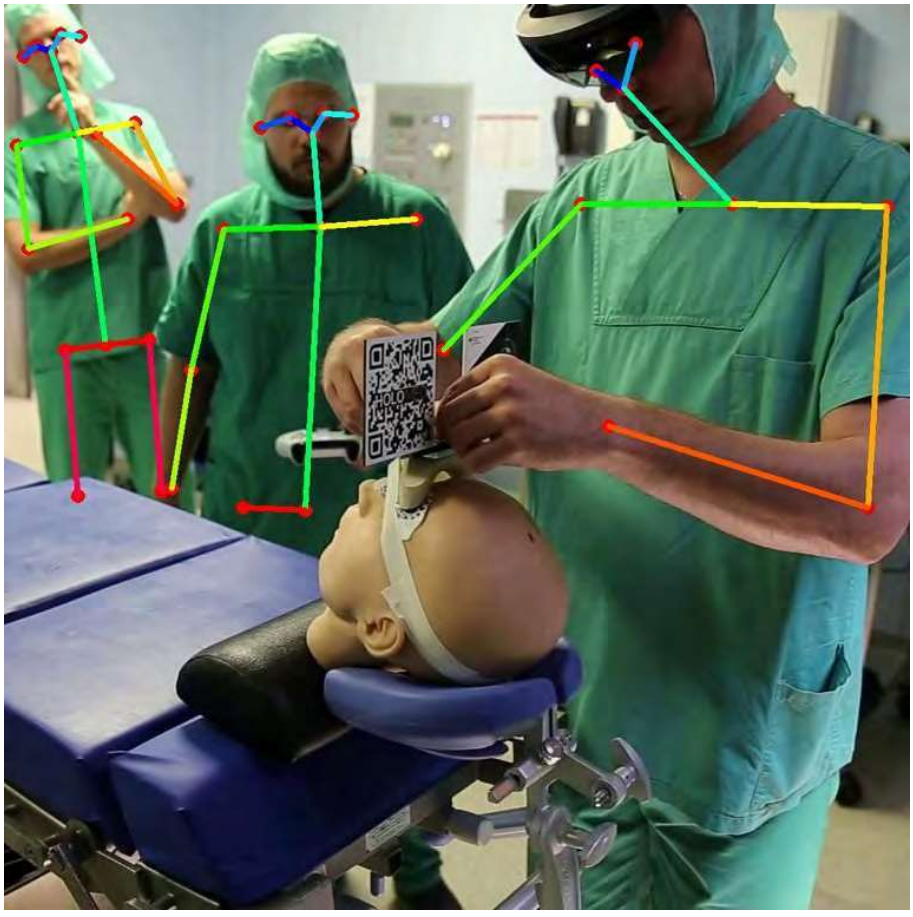


Fig. 11. Results of the medical staff tracking.

4 Discussion

As shown in the evaluation, our approach succeeds in detecting the patient in an automated two-stage process with an accuracy between 3 and 7 mm. The coarse position is determined by using a YOLO v3 net. The results under normal OR conditions are very satisfying. The solution performance drops strongly under bright illumination conditions. This is due to large flares that occur on the phantom as it is made of plastic or silicone. However, these effects do not occur on human skin. The advantage of our system is that the detection is performed on all four Kinect RGB streams enable different views on the operation area. Unfavourable illumination conditions normally don't occur on all of these streams. Therefore a robust detection is still possible. In the future the datasets will be expanded with samples with strong illumination conditions.

The following surface matching of the head yields good results and a robust and precise detection of the patient. Most important is a good preprocessing of the CT data and the recorded point cloud of the search area, as described in the methods. The algorithm does not manage to find a result if there are larger holes in the point clouds or if the normals are not calculated correctly. Additionally, challenges that have to be considered include skin deformities and noisy CT data. The silicone skin is not fixed to the skull (as human skin is), which leads to changes in position, some of which are greater than 1 cm. Also the processing time of 7 minutes is quite long and must be optimized in the future. The processing time may be shortened by reducing the size of the point clouds. However, in this case the matching results may also become worse.

Catheter tracking [7] yielded good results, despite the challenging task of segmenting a very thin (2.5 mm) and deformable object. Additionally, a 3D estimation of the catheter was implemented. The results showed that in many cases over 90% of the catheter can be estimated correctly. However, these results strongly depend on the orientation and the quality of the depth stream. Using higher quality sensors could improve the detection results.

For tracking of the medical staff OpenPose as a ready-to-use people detection algorithm was used and integrated into ROS. The library produces very good results, despite medical staff wearing surgical clothing.

5 Conclusion

In this work the integration of augmented reality into the digital operating room OP:Sense is demonstrated. This makes it possible to expand the capabilities of current AR glasses. The system can determine the precise patient's position by implementing a two-stage process. First a YOLO v3 net is used to coarsly detect the patient to reduce the search area. In a second subsequent step a two-stage surface matching process is implemented for refined detection. This approach allows for precise location of the patient's head for later tracking.

Further, a FRNN-based solution to track the surgical instruments in the OR was implemented and demonstrated on a thin neurosurgical catheter for ventricular punctures. Additionally, OpenPose was integrated into the digital OR to track the surgical personnel. The presented solution will enable the system to react to the current situation in the operating room and is the base for an integration into the surgical workflow.

6 Acknowledgement

This work was partially funded by the Federal Ministry of Education and Research within the project 'HoloMed - Context-sensitive support of a surgeon in the operating room by Augmented Reality'. The authors acknowledge the expertise provided by neurosurgeons from the Department of Neurosurgery at the University Hospital Ulm/Günzburg.

References

1. Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: realtime multi-person 2d pose estimation using part affinity fields. arXiv preprint arXiv:1812.08008, 2018.
2. Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd, volume 96, pages 226–231, 1996.
3. Christian Kunz, Pit Henrich, Paul Maria Scheikl, Michal Hlavac, Max Schneider, Woern Heinz, Bjoern Hein, and Franziska Mathis-Ullrich. Fast volumetric auto-segmentation of head ct images in emergency situations for ventricular punctures. In CURAC, Reutlingen, Germany, September, 2019.
4. Christian Kunz, Michal Hlavac, Max Schneider, Bjoern Hein, David Puljiz, and Steffen Peikert. A system for augmented reality guided ventricular puncture using a hololens: Design, implementation and initial evaluation. In CURAC, Leipzig, Germany, September, 2018.
5. Holger Mönnich, Heinz Wörn, and Daniel Stein. Op sense—a robotic research platform for telemanipulated and automatic computer assisted surgery. In AMC 2012, pages 1–6. IEEE, 2012.
6. Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. arXiv, 2018.
7. Fabian Reister, Christian Kunz, Max Schneider, and Torsten Kroeger. Deep learning based 3d pose estimation of surgical tools using a rgb-d camera at the example of a catheter for ventricular puncture. In CURAC, Leipzig, September, 2018.
8. Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3d registration. In 2009 IEEE international conference on robotics and automation, pages 3212–3217. IEEE, 2009.