Investigations on 3D reconstruction of bones in surgery using a handheld trinocular camera system

Arne Schierbaum [1], Tobias Neiss-Theuerkauff [2], Thomas Luhmann [1], Frank Wallhoff [2], Till Sieberth [1]

[1] Institute for Applied Photogrammetry and Geoinformatics, Jade University of Applied Sciences, Oldenburg, Germany
(arne.schierbaum, luhmann, till.sieberth)@jade-hs.de
[2] Institute for Technical Assistive Systems, Jade University of Applied Sciences, Oldenburg, Germany
(neiss-theuerkauff, frank.wallhoff)@jade-hs.de

Technical Commission II

Keywords: SLAM, trinocular camera system, segmentation, COLMAP, surgery, knee arthroplasty

Abstract

Knee arthroplasty benefits significantly from computer-assisted navigation, which improves the accuracy of prosthesis placement. However, current methods require invasive optical locators to track the position of the knee, which carries risks such as infection and prolonged healing times. To address these limitations, this work uses markerless trinocular SLAM to achieve accurate 3D reconstruction of the knee during surgery. The approach integrates SuperGlue for robust feature matching and incorporates segmentation to mask the knee, improving reconstruction accuracy despite challenges such as low-texture surfaces, reflections and spotlight illumination. The accuracy of the handheld trinocular camera system is evaluated under dynamic conditions, simulating camera movement during surgery to ensure accurate reconstruction during real-time surgery. In addition, a robot-guided dataset will be used to assess the repeatability and robustness of the SLAM approach. This research focuses on positional accuracy in motion and aims to advance real-time, non-invasive navigation solutions for knee arthroplasty, contributing to safer and more efficient surgical outcomes.

1. Introduction

In knee arthroplasty, computer-assisted navigation is an important method for increasing the precision of the placement of the artificial knee (Moret and Hirschmann, 2021). Approximately 29% of operations in knee arthroplasty are already supported by assistive, partially navigated systems (Rath et al. 2011). In general, the navigation procedure is used to support the surgeon in placing a saw block, as guide for the subsequent bone cut. However, currently it is necessary to drill optical locators into the bone in order to track the knee during the operation (Figure 1.). The locators allow the knee to be determined in relation to the patient's leg axis, which is defined by the center of the hip and ankle joint. The disadvantage of this method is that the drilled holes increase the risk of infection and also prolong the healing process of the patient (Stübig et al. 2020). Hence avoiding the drilling procedure is the main motivation for the presented research.

A markerless navigation procedure could allow to transfer the pre-planned position of the sawblock into the surgery area without the need for drilling into the patient's bone. To achieve these two main objectives, need to be accomplished. First the localisation of the knee and leg axis, and secondly the visualisation of the cut position on the leg. While the latter is to be achieved using augmented reality, the localisation procedure can be performed using photogrammetric methods. A proof-of-concept study by Hu et al. (2021) suggests that marker-free knee surgery using photogrammetric techniques offers promising results. However, conventional SLAM (simultaneous localisation and mapping) methods reach their limits in surgical applications as they do not sufficiently meet the specific requirements, as shown by Kahmen et al. (2020). There, a trinocular camera system with visual odometry is proposed to reconstruct the knee surface. The potential is shown that the knee can be successfully reconstructed with a stable camera system. Our aim is that during surgery, the visible area of the femur and tibia is digitized using photogrammetry. Then, this resulting model is aligned with pre-operative CT scans to establish the legs

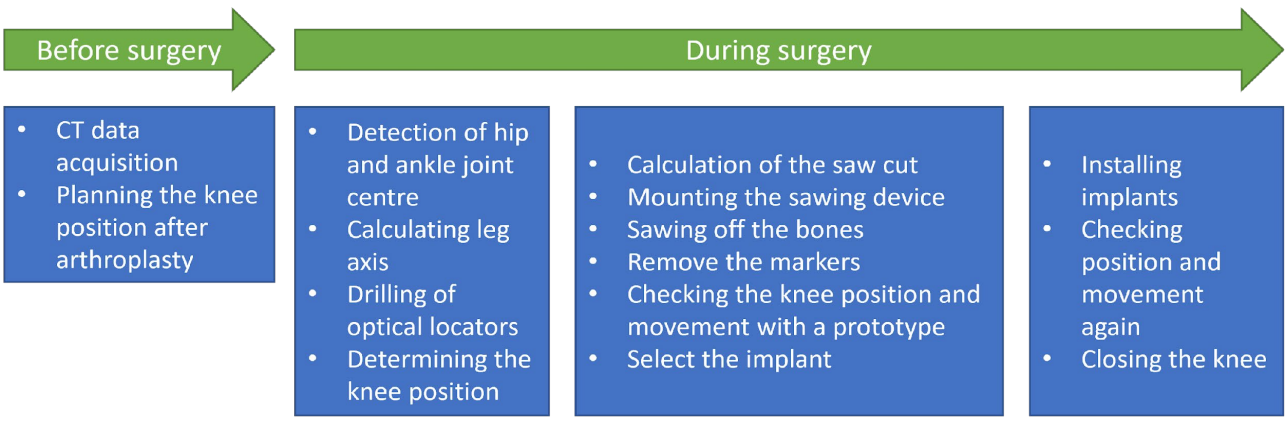

Figure 1. Current stages of navigated knee surgery

axis. After successfully aligning pre-operative CT with planned saw position, it is possible to transfer the saw position onto the bone. Therefore, the visible areas of the bones should be reconstructed as completely as possible, as this leads to better orientation of the knee itself. This presents challenges as the bones have less texture and the wet surface causes reflections, which are intensified by the punctual illumination. Additional due to blood or tissue covering the surface. Another challenge is presented in the movement of the knee: This causes femur and tibia to move simultaneously and independently.

The basic idea of the proposed markerless trinocular SLAM system for knee arthroplasty is to first acquire detailed 3D images of the knee joint in a static state. In this first step, the surface of the knee is captured from different angles to create an accurate model. Once the model has been created, the system could then move into the active phase of the surgery, using it to navigate the procedure in real time. The reconstructed knee model would act as a reference point, allowing the surgeon to track and control movements with high precision. This approach would enable non-invasive navigation during knee arthroplasty without relying on physical markers.

We aim to realise a marker free navigation by generating a highly accurate 3D reconstructions of the knee surface in real time - despite the low texture and interfering influences such as reflections, shading or occlusions.

2. Trinocular SLAM based on SuperGlue

The challenges listed lead to the implementation of a trinocular SLAM based approach based on the work of COLMAP SLAM (Morelli et al., 2023). To orient the image sequence in trinocular SLAM we cannot use the entire image content. This is because the knee is moving during surgery and therefore the configuration of femur and tibia and the surrounding tissue is changing. Neiss-Theuerkauff et al. (2024) presents an AI-based segmentation for femur and tibia which we integrate to mask the respective bones in the images for further processing. On these masked images we used the AI-based feature matching algorithm SuperGlue (Sarlin et al. 2020). Based on the matched features a precise relative orientation can be calculated, which is crucial for the knee reconstruction (Figure 2.). The movement of the trinocular camera system multiple allows us to acquire images in sequence during multiple epochs from different points of view. Each image triplet is matched to the triplet of the two previous epochs of the sequence. We use COLMAP to align the images in a bundle during acquisition (Schönberger and Frahm, 2016). COLMAP calculates the bundle adjustment without considering a scale, therefore the pre-calibrated camera system defines the scale of the 3D reconstruction.

2.1 Handheld trinocular camera system

There are many challenges to achieve a markerless 3D reconstruction of the knee. Kahmen et al. (2020) designed a trinocular camera system and discussed the advantages in medical surgery. Due to the limited space in the surgery room, a small handheld measurement system is required. In our application we use a similar trinocular system as a handheld device (Figure 3). The trinocular system consists of two monochromatic cameras and one RGB camera, as developed by Kahmen et al. (2020). The three cameras are arranged in an equilateral triangle with a side length of 5 cm and tilted inwards rotated by about 2 degrees. This results in large image overlay at a distance of approximately 60 cm. The camera system has a handle at the bottom for free movement of the device.

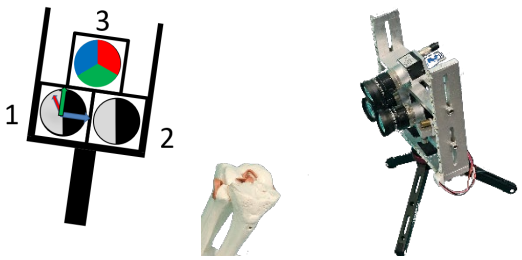

Figure 3. Setup of the trinocular camera (left, (Kahmen et al. 2020)) and when aligned with an anatomical knee model (right, (Schierbaum et al. 2024))

Before the camera system is used, it is calibrated using a test 3D field. The relative orientation and the interior parameters are determined. For this propose, approximate values are first determined using AICON 3D Studio (Version 12.0 by AICON 3D Systems, Germany). AxOri (by AXIOS 3D Services,

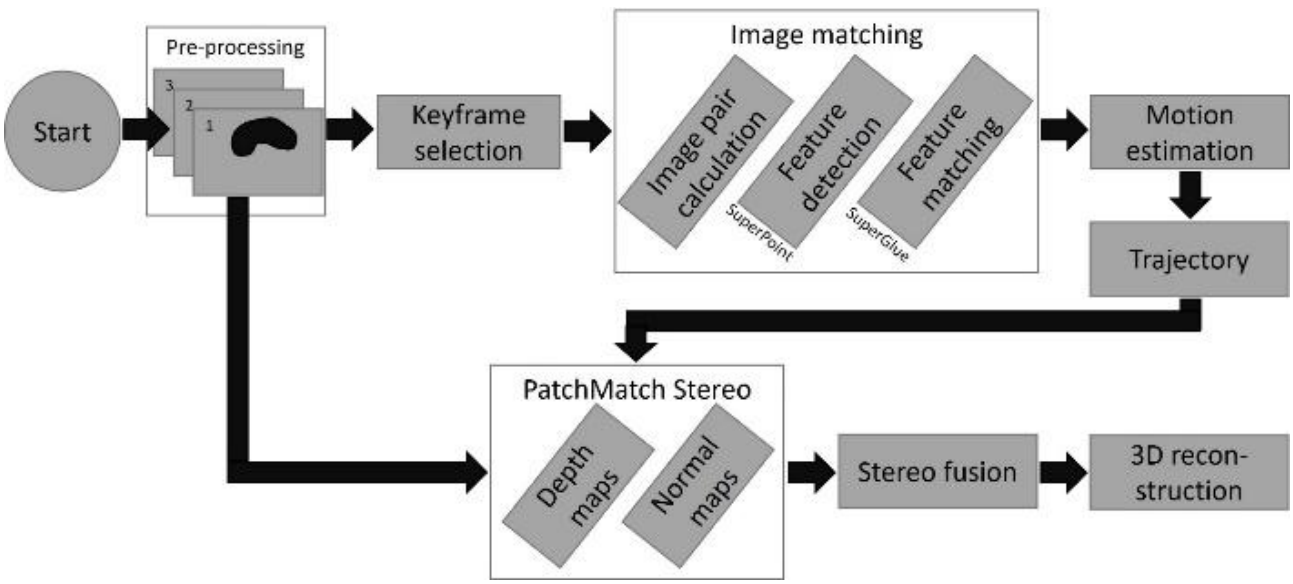

Figure 2. Process of the trinocular SLAM with COLMAP and SuperGlue (Schierbaum et al., 2024)

Germany) is then used to calculate a bundle adjustment with fixed relative orientation constraint in the camera system. We place the coordinate origin in the projection centre of the first camera. The interior orientation parameters must be transformed from metric to pixel-based due to COLMAP requirements. The conversion equations are given in Luhmann (2023).

2.2 Image pre-processing

The first step after image acquisition is to pre-process the images. To orient the image sequence in trinocular SLAM we cannot use the entire image content because the knee is moving during surgery. The dynamic scene challenges the SLAM. Therefore, a segmentation is used to mask the regions where the bones are located (Neiss-Theuerkauff et al. 2024). The masking is also used to filter out the images where no bones are visible. A second filter method is the keyframe selection by Morelli et al. (2023), which checks the image content for innovation. In our case we adapt their method to perform the keyframe selection on images of the first camera. We use feature matching, based on SuperGlue (Sarlin et al. 2020), to control the overlapping regions between successive images in the recorded sequence. The keyframe is accepted for the following workflow if the matching has at least 20 matches.

2.3 Trajectory

To reconstruct an accurate 3D model of the knee surface during surgery we used a feature-based method to predict the movement of our handheld camera system. This involves orienting the images of the trinocular system in a local coordinate system.

2.3.1 Image pairs

Before starting the feature matching between two images of our sequences we have to define the image pairs. The selection of the image pairs depends on the one hand on the highest possible accuracy of the image orientation. On the other hand, the aim is to calculate the 3D model in real time. From these two points of view, we investigate different methods to generate the image pairs. For each new image triplet, we obtain new image pairs. The pairs can be ordered in three groups and are shown in Figure 4.

Pairs of images between:

1. same epoch (black arrows)
2. current epoch and previous epoch from the same camera (green arrows)
3. current epoch and previous epoch from the same camera (blue arrows)

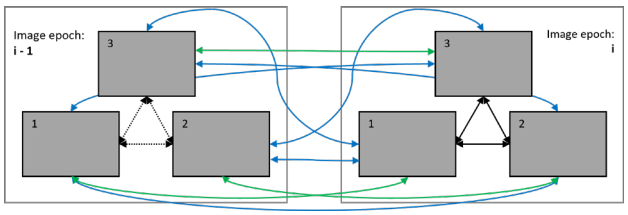

Figure 4. Schematic representation of image pairs between two consecutive image epochs

We build our investigation methods from these three groups. The first method includes all image pairs of the two first groups (black and green arrows). The second method includes all image pairs of the first and the third group (black and blue arrows). The third method includes all groups (black, green and blue arrows). In addition, in all methods we take into account the image pairs that we can generate with the images of the second preceding epoch *(i-2)*. All different methods are listed in Table 1. The last column shows the new image pairs that will be added in the next epoch *i*. As the image pairs of first group already exist, they are not determined again (dotted black arrows in Figure 4.).

Method	Group			Epoch		Pairs / Epoch
	1	2	3	i - 1	i - 2	
1	●	●		●	●	9
2	●		●	●	●	15
3	●	●	●	●	●	21

Table 1. Different methods to generate image pairs in the trinocular system

2.3.2 Image matching

The formed image pairs are the basis for following feature-based image matching. In our workflow we use a two-step method consisting of SuperPoint (DeTone et al. 2018) and SuperGlue (Sarlin et al., 2020). SuperPoint is a CNN-based feature detector, which recognises and describes features in an end-to-end approach. The architecture uses synthetically generated point labels for training, making it robust to various transformations without manually labelled data. It combines feature detection and description in one network, enabling close alignment between the two, improving recognition and stability. SuperPoint creates an even distribution of points in the image, which supports a more informative and balanced feature representation (DeTone et al. 2018). In our application, the search area is limited by features due to masking. Therefore, the area is small in which many high-quality features can be found and the bones have homogeneous surfaces. This results in a non-maximum-suppression value of 4 pixel and a keypoint threshold of 0.001. The two values regulate the quality and quantity of the features.

After detection, the graph neural network SuperGlue match the features between the defined images pairs. It models the features of both images as graphs and uses self- and cross-attention mechanisms to incorporate contextual information and relationships between features. The features are iteratively enriched in a common feature space, which enables robust and consistent matches even with significant perspective and illumination differences.

2.3.3 Motion estimation

The matches and their corresponding features, as well as their descriptors, are transferred to the COLMAP database. We also load the intrinsic and relative orientation of our trinocular camera system in COLMAP. In a first batch of images the COLMAP *mapper* starts with an initial pose estimation and the 3D points between the first image pair. New images in this batch are added to be registered in the COLMAP model. From the registered images, features can be added to the sparse point cloud by triangulation. The bundle adjustment iteratively optimises the reprojection error, increasing the accuracy of the camera position and orientation, as well as the sparse point cloud (Schönberger and Frahm 2016).

The bundle adjustment does not consider the relative orientation of the camera system, which results in a scale-free reconstruction. Therefore, we scale the reconstruction with the first image triplet. To reduce the uncertainty of the image triplet a bundle adjustment with stable camera system is performed.

2.4 3D reconstruction

For the 3D reconstruction using Multi-View Stereo (MVS), source images are required for which depth and normal maps are

created. For each of these source images, reference images are then searched which fulfil two prerequisites. Therefore, the trajectory and the correspondence sparse point cloud form the basis for the 3d reconstruction of the bones. These are used to estimate the overlapping images. There must also be an angle of between 5° and 20° between the source and reference images. If the angle is too small, the geometry is not sufficient for an appropriate depth estimate. If the angle is too large, the overlapping image content becomes too small, which leads to a poor 3D reconstruction (Schönberger et al. 2016).

In our workflow, we defined at least six source image triplets. These image triples show the masked knee from different directions to ensure a complete reconstruction. Once the reference images have been defined for each source image, the PatchMatch Stereo (Zheng et al. 2014) process begins to compute dense depth maps. Each pixel in the source images goes through several steps to obtain an accurate depth estimate. First, each pixel is assigned a random depth hypothesis, which is then optimised across neighbouring pixels. By propagating and comparing the depth values between neighbouring pixels, better estimates can be found. The depth values are then randomly varied to find locally optimal depth solutions. This process is performed iteratively until stable and consistent depth maps are computed for each source image (Schönberger et al. 2016).

The depth maps are then merged. As several depth maps from different reference views were calculated for each source image, these maps are now combined into a dense 3D point cloud. Firstly, unreliable depth values are filtered in order to retain only the consistent and stable depth values. This adjusted depth information is then localised by triangulation in a common 3D scene. Care is taken to ensure that each point has only one unique 3D position, even if it is visible in multiple depth maps. This fusion produces a dense point cloud of the knee (Schönberger et al. 2016).

3. Experimental setup

An experimental setup was used to evaluate the trinocular SLAM. For this purpose, an artificial knee joint was placed within a reference frame and recorded by the trinocular camera system. The frame provides a reference data to determining the orientation of the trinocular system (Kahmen et al. 2020). The knee joint and the frame were placed on a table covered with blue surgical cloth. The setup was illuminated with two spotlights that shine on the knee from opposite directions to ensure shadow free lighting. In a first dataset, the hand-held trinocular camera system is moved in a circle over the artificial knee. The camera system is moved in such a way that there is no rotation around the visual axis. During the motion two images are taken per second. The dataset includes a total of 30 image triples. The ground truth of the exterior orientation is shown in Figure 5. It has been equalised in a bundle adjustment using the circular markers, taking into account the fixed relative orientation between the cameras.

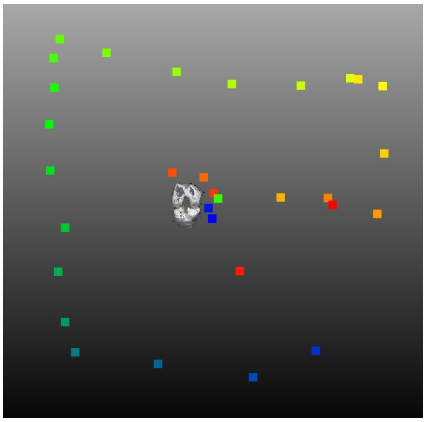

Figure 5. Ground truth trajectory of the handheld dataset for the first camera. First epoch (blue) to last epoch (red)

A second dataset is created using the robot “Fanuc CR-7iA/L”. The idea of using the robot is that we cannot give any information about the precision of our trinocular SLAM. The robot offers the possibility of acquiring images several times in comparable situations. The trinocular camera was attached to robot as presented in Figure 6. The camera was then moved along a predefined path. As stated by the manufacturer the robot has a repeat accuracy of 0.01 mm. Setting the motion speed to 80 % of the maximum speed, the robot achieved a speed of 800 mm/s. This corresponds approximately to the movement speed when the camera system is moved slowly hand. In addition to the linear motion, we predefined stopping positions to allow for stable image acquisition.

Figure 6. Experimental setup with the robot and the artificial knee joint surrounded by the reference frame with two spotlights from opposite sides

In this second dataset, we distinguish between two recording configurations. In a first dataset the knee is recorded together with the reference frame. For a second dataset the reference frame is removed and only the knee is captured by the trinocular camera system. The knee remains in the same position during all recordings. The image sequences consist of each 59 image triples and are therefore twice as large as the first dataset. The first sequences with a reference frame should serve as a ground truth for the second sequence without the frame. However, as these differ greatly from one another and the robot does not always move to the positions at the same speed, there is only a reference for the image triples when the robot is stationary. As the image triples between the fixed positions are included in the evaluation, the influence of the camera movement can still be determined at

the stable positions. Figure 7. shows the three image sequences with the reference data. Between the predefined positions (no. 1-9) you can see the different deviating locations of the images in the three reference datasets.

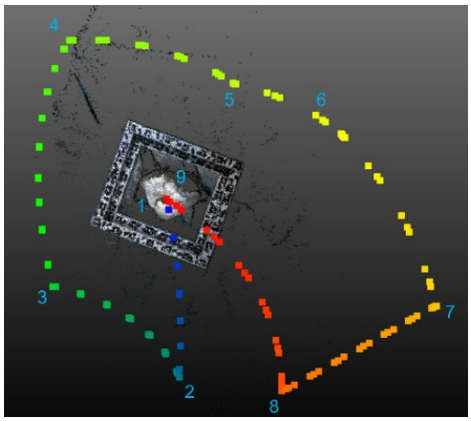

Figure 7. Ground truth trajectory of the triple robot dataset for the first camera. First epoch (blue) to last epoch (red) The numbers from 1 to 9 show the fixed positions of the robot (epochs: 0, 9, 17, 21, 28, 34, 43, 51 and 58).

The two experimental setups result in the following datasets listed in Table 2. We analyse the data sets both with and without masks, in which case the reference frame is also visible in the images. Dataset 1 is the reference for the datasets 2 and dataset 4 is the reference for dataset 5 to 8 in robot guided case. In the reference datasets, the orientation of the trinocular system is determined with an accuracy of better than 0.05 mm in translation and 0.005° in rotation, which is well above the expected accuracy for our SLAM.

Set	Robot or handheld	Knee masked	Orientation technique
1	handheld	--	Bundle adjustment
2	handheld	No	Trinocular SLAM
3	handheld	Yes	Trinocular SLAM
4	Robot with frame	--	Bundle adjustment
5	Robot with frame	No	Trinocular SLAM
6	Robot with frame	Yes	Trinocular SLAM
7	Robot	No	Trinocular SLAM
8	Robot	Yes	Trinocular SLAM

Table 2. Datasets that are used in the work. Datasets 1 and 4 form the reference datasets of the two experimental setups. The reference for datasets 7 and 8 is created for the stable positions by dataset 4.

4. Accuracy of the trinocular SLAM

To compare the absolute coordinates and orientations of the camera trajectory, it is necessary to equate the origin and the orientation of the coordinate system (Kahmen et al. (2020), Therefore, the local coordinate system of the trajectory is transformed into the projection centre of the camera at the first epoch and the z-axis is transformed into the optical axis.

4.1 Accuracy dependence of image pairs

As we introduced our workflow, we presented the various methods for creating image pairs. Starting with the handheld dataset, we analyse the deviations compared to our reference. The dataset is calculated once with the masked images, where the camera orientation is only determined via the knee surface, and in a second step without masking, so that the markers on the reference frame can be used. Table 3. shows the results for these tests, where the mean Euclidean distance to the reference (Set 1) is given and the mean angular deviation is listed.

No.	Set	Method	RMS 3D [mm]	Max 3D [mm]	RMS [°]	Max [°]
1	2	1	7.23	13.39	0.35	0.77
2	3	1	6.32	11.48	0.79	1.50
3	2	2	6.17	11.87	0.36	0.82
4	3	2	calculation interrupt			
5	2	3	6.26	11.95	0.37	0.83
6	3	3	6.08	11.61	0.78	1.54

Table 3. Results of the investigation into various methods for generating image pairs.

For the handheld dataset, trinocular SLAM archives around a mean Euclidean distance of 6 mm and more. The accuracy of the camera orientation shows no significant difference between the different methods. Only two evaluations stand out. First, the calculation of the orientation with method 2 could not be completed for the masked image sequence in which no sufficient feature matches were found in the images. Second, the accuracy of result 1 is lower than the other results.

4.2 Accuracy robot-guided image sequences

In the following investigation, the repeatability of the accuracy of a moving image recording is to be analysed. For this purpose, we use the dataset in which the robot holds the trinocular system to be able to analyse three image sequences independently of each other. Following the surprising results from the previous study, we also analyse the images with and without masking. The results are shown in Table 4.

No.	Set	Method	RMS 3D [mm]	Max 3D [mm]	RMS [°]	Max [°]
7	5.1	3	9.50	18.85	0.39	0.76
8	6.1	3	4.90	9.10	0.38	0.75
9	5.2	3	9.38	18.97	0.38	0.72
10	6.2	3	4.88	8.85	0.36	0.65
11	5.3	3	8.86	17.98	0.38	0.75
12	6.3	3	4.27	7.27	0.54	1.15

Table 4. Accuracy of robot-held images sequences

While the analysis with and without masking produced comparable deviations in the previous study, the situation is different with the robot dataset. The average distance is approximately twice as good for the masked images. And the overall accuracy is also higher than in the handheld dataset. In addition, the absolute distances between two epochs are smaller, so that the images are similar and the features can therefore be assigned better. It is noteworthy that the trajectory has twice as many epochs as the handheld dataset and still delivers the better results.

4.3 Accuracy frameless image sequence

A third investigation into the orientation of the trinocular camera system is realised using the robot data set. For this purpose, we only refer to the specific image epochs in which the robot travelled to the predefined position.
Based on these results, we analyse the image sequence without a reference frame and also refer to the nine positions shown in

Figure 7. From these image triples, we now calculate the mean deviation at these positions for our datasets.

No.	Set	Method	RMS 3D [mm]	Max 3D [mm]	RMS [°]	Max [°]
8	6.1	3	5.06	8.89	0.30	0.57
10	6.2	3	4.36	8.62	0.30	0.51
12	6.3	3	4.09	5.33	0.49	0.78
13	7	3	8.34	16.13	0.59	1.11
14	8	3	6.95	13.81	0.57	1.28

Table 5. Accuracy of robot-held images sequences at positions 1 to 9.

The average deviation is comparable for the predefined positions of the robot held dataset (6.1, 6.2, 6.3) and for the whole trajectory. However, when analysing the data set without a reference frame, the deviation is higher at 6.95 mm when the images are masked. Without masking, an accuracy of 8.34 mm is obtained, which is in the accuracy range of the other image sequences where the reference frame is in the image.

4.4 Orientation of the 3D reconstruction

The orientation of the camera system is initially an intermediate step on the way to the goal of orientating the knee joint in relation to the leg axis. This is why knowledge about the orientation accuracy of the 3D reconstruction is important. To do this, we use the robot dataset and use COLMAP to calculate a dense point cloud from the image triples taken at the defined position. This is to ensure that the dense point cloud is only created from images that were taken from the same positions. In order to obtain a ground truth, we generate a dense point cloud, also using COLMAP, based on the camera orientation determined using bundle adjustment. To analyse our trinocular SLAM, the point clouds from our SLAM and the ground truth are then matched using an ICP algorithm in the CloudCompare (Version 2.12.4) software.

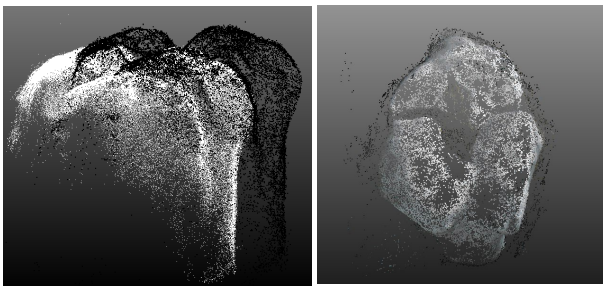

Figure 8. Reference point cloud (white) and the point cloud created by our trinocular SLAM (black) from the 9 image triplets (left) and a top view of the point cloud using trinocular SLAM.

The point clouds of the three datasets with reference frames around the artificial knee have absolute deviations between 14 mm and 22 mm. Figure 8. on the left shows the point cloud of dataset 6.3. The deviation is 21.3 mm. In the orientation of the knee joint, the deviation is less than two degrees around the x-axis. This shows that the deviations in the camera orientation are transferred to the position of the reconstruction and there are larger deviations, particularly in the translation of the knee joint.

5. Discussion

The results of this study provide valuable insight into the performance of trinocular SLAM for knee reconstruction. As expected, the accuracy in this study was lower compared to previous works such as Kahmen et al. (2020) and Schierbaum et al. (2024). Several factors contribute to this lower accuracy. First, the camera movement in our dataset led to lower image quality, which in turn negatively affected the accuracy of the image measurements. Blurred images can introduce errors into the reconstruction process. In addition, the camera was positioned approximately 60 cm away from the artificial knee joint, as opposed to the 30 cm distance used in Kahmen et al. (2020). This increased distance reduces the resolution of the images by a factor of 4, which leads to a decrease in detail capture and ultimately affects the quality of the reconstruction. In addition, the shading of the knee joint further complicates the process as it can obscure surface features and prevent effective feature extraction.

One observation is that the results obtained with reference frames did not show higher accuracy than those without. In previous studies, reference frames have typically led to better results because it is easier to match corresponding features across images. However, in this study, the expected improvements from reference frames were not observed. This could be due to the specific settings of SuperGlue, which are optimised for textureless surfaces, resulting in fewer features matching on the reference frame.

Another interesting finding is that the dataset of robot-held images produced better results than the hand-held dataset. The robot's movements were more homogeneous and repeatable, resulting in fewer outliers and more consistent data. In contrast, handheld data collection often introduces more variability, which can reduce overall accuracy. This highlights the importance of stable and controlled camera movement for high accuracy reconstructions in these applications.

Despite these challenges, the trinocular SLAM system showed a promising ability to produce point clouds without large holes or significant gaps. Such problems were more common in previous studies using similar exposure settings, suggesting that the system is able to deal effectively with variations in reflections. These variations, which are influenced by the movement of the camera, are likely to contribute to a more complete reconstruction of the knee surface as reflections do not always occur in the same locations.

It is important to note that for the purposes of this study, the knee was treated as a stable object. This assumption allowed a simplified assessment of the accuracy of the system, but in a real surgical context the knee joint itself would be subject to dynamic movements due to operational actions. These movements introduce additional complexity that will affect the accuracy of the reconstruction and the overall effectiveness of the navigation system.

6. Conclusions and outlook

This research demonstrates the potential of markerless trinocular SLAM for non-invasive navigation in knee arthroplasty, although the accuracies achieved still have some weaknesses. Camera motion, combined with challenges such as the low texture of the knee surface, has prevented the achievement of sub-millimetre accuracies. This is mainly due to factors such as camera movement, the increased distance to the knee joint and

the difficulty in capturing detailed information from a surface with poor texture. In particular, the increased distance compared to previous studies affects the resolution of the captured images and reduces the accuracy of the reconstructions.

The results from the robotic dataset, which had more homogeneous and repeatable movements, showed better consistency and accuracy. These results provide a solid basis for further investigation under controlled conditions. This highlights the importance of controlled camera movements to ensure high quality reconstructions and improve the accuracy of 3D modelling. To further improve accuracy, adjustments to the camera settings are needed to improve image quality.

Another important step for future work is to perform tests with more realistic data. The current work is primarily based on artificial knee data sets, while the challenges associated with real surgical environments, including dynamic knee movements and the presence of tissue or blood, have not been fully considered. Further testing with real data will help to identify additional challenges and potential areas for improvement. In addition, it will be necessary to investigate how the reconstructed knee can be aligned with the leg axis.

In conclusion, the trinocular SLAM system is a promising approach for non-invasive navigation in knee arthroplasty, but improvements are needed to enable marker-free navigation. With further optimisation of the system, it could potentially contribute to intraoperative knee registration in the future.

Acknowledgement

The research work presented was funded by the Bundesministerium für Bildung und Forschung (BMBF). Many thanks for the co-funded support from Aesculap and AXIOS 3D Services as well as the insights and expertise of the PIUS Hospital (Oldenburg).

References

DeTone, D., Malisiewicz, T. & Rabinovich, A. (2018): SuperPoint: Self-Supervised Interest Point Detection and Description. *CVPR 2018 Deep Learning for Visual SLAM Workshop,* IEEE, 224-236.

Hu, X., Liu, H. & Baena, F.R.Y. 2021: Markerless Navigation System for Orthopaedic Knee Surgery: A Proof of Concept Study. *IEEE Access*, 9, 64708-64718. doi.org/10.1109/ACCESS.2021.3075628

Kahmen, O., Haase, N. & Luhmann, T., 2020: Orientation of point clouds for complex surfaces in medical surgery using trinocular visual odometry and stereo ORB-SLAM2. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*., doi.org/10.5194/isprs-archives-XLIII-B2-2020-35-2020

Luhmann, T. 2023: *Nahbereichsphotogrammetrie. Grundlagen - Methoden - Beispiele.* Wichmann, Berlin/Offenbach.

Morelli, L., Ioli, F., Beber, R., Menna, F., Remondino, F. & Vitti, A., 2023: COLMAP-SLAM: A framework for visual odometry. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*., 317-324. doi.org/10.5194/isprs-archives-XLVIII-1-W1-2023-317-2020

Moret, C.S. & Hirschmann, M.T., 2021: Navigation und Robotik in der Knieendoprothetik. *Arthroskopie*, 34 (5), 351-357.

Neiss-Theuerkauff, T., Schierbaum, A., Luhmann, T., Sieberth, T., Wallhoff, F., 2024: Semantic Bone Structure Segmentation in 2D Image Data: Towards Total Knee Arthroplasty. *Artificial Intelligence XLI - 44rd SGAI International Conference on Artificial Intelligence, AI 2024, Cambridge, UK, December 17-19, 2024, Proceedings. Lecture Notes in Computer Science (LNCS), Springer 2024 (to appear)*

Sarlin, P.-E., DeTone, D., Malisiewicz, T. & Rabinovich, A., 2020: SuperGlue: Learning Feature Matching with Graph Neural Networks. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* IEEE. 4938-4947 doi.org/10.1109/cvpr42600.2020.00499

Schierbaum, A., Neiß-Theuerkauff, T., Wallhoff, F., Sieberth, T., Luhmann, T., 2024: Untersuchungen zu einem KI-basierten SLAM-verfahren für ein trinokluares Kamerasystem zur 3D-Erfassung der Knieoberfläche. *Photogrammetrie Laserscanning Optische 3D-Messtechnik*

Rath, B., Springorum, H.-R., Beckmann, J., Schaumburger, J., Tingart, M., Grifka, J. & Lü-ring, C. 2011: Aktueller Stellenwert der Navigation in der Knieendoprothetik in ortho-pädischen und unfallchirurgischen Kliniken in Deutschland, *Zeitschrift fur Orthopadie und Unfallchirurgie*, 149 (2), 173–177

Schönberger, J. L., Frahm, J.-F., 2016. Structure-from-Motion Revisited, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2016, pp. 4104-4113 doi.org/10.1109/CVPR.2016.445

Schönberger, J.L., Zheng, E., Frahm, J.-M. & Pollefeys, M., 2016: Pixelwise View Selection for Unstructured Multi-View Stereo. *Computer vision - ECCV 2016*, 501-518. doi.org/10.1007/978-3-319-46487-9_31

Stübig, T., Windhagen, H., Krettek, C. & Ettinger, M., 2020: Computer-Assisted Orthopedic and Trauma Surgery. *Deutsches Ärzteblatt international, 117 (47),* 793-800. doi.org/10.3238/arztebl.2020.0793

Zheng, E., Dunn, E., Jojic, V. & Frahm, J.-M. 2014: PatchMatch Based Joint View Selec-tion and Depthmap Estimation. *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 1510-1517.