Trinocular Multi-Object 3D Reconstruction in Camera-Simulating Virtual Environments for Knee Arthroplasty

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

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

Keywords: 3D Reconstruction, Trinocular Camera System, Synthetic Data, COLMAP, Surgery

Abstract

In knee arthroplasty, computer-assisted navigation enhances the accuracy of prosthesis placement. However, current methods rely on invasively drilled locators to track the knee position during surgery, prolonging the healing process. For this reason, research is focused on markerless approaches capable of determining knee orientation and transferring preoperative planning into the surgical environment. This work presents a trinocular multi-object 3D reconstruction system designed for intraoperative acquisition of the knee surface, providing a foundation for marker less navigation. Due to the scarcity of real surgical data with ground truth, a synthetic dataset was created using Blender to simulate optical image acquisition of a virtual knee model under controlled camera and lighting conditions. The dataset enables a systematic evaluation of how camera motion and viewpoint affect pose estimation and 3D reconstruction accuracy. The results demonstrate that moderate camera deflection between 15° and 25° achieve the best balance between accurate camera pose estimation and surface reconstruction quality. The work confirms the potential of trinocular SLAM for robust bone surface tracking while also identifying the limitations of synthetic data, such as the absence of real-world visual variability. These results form the basis for future work on 3D reconstruction during dynamic knee movements and their tracking, as well as on the integration of markerless optical navigation systems into surgery.

1. Introduction

Knee arthroplasty is a frequently performed surgical procedure in which computer-assisted and partially robot-guided systems are increasingly used to enhance precision (Moret and Hirschmann, 2021). A key component of such systems is surgical navigation, which transfers preoperative planning to the intraoperative situation and assists the surgeon in accurately aligning the saw block under dynamic conditions. Conventional navigation systems rely on optical markers rigidly fixed to femur and tibia, allowing continuous tracking of the knee. However, these invasive markers require drilling into the bone, which can prolong healing and increase the risk of infection (Stübig et al., 2020).

To enable markerless navigation in the future, the knee surface must be captured to register preoperative CT data with the intraoperative 3D reconstruction of the bones. This would allow the transfer of planning data into the surgery. Continuous tracking of the individual bones enables real-time navigation, which requires highly accurate segmentation of the bones. As part of the ASKAR3D project, this work focuses on the development of fundamental methods for evaluating dynamic scenes using approaches related to Simultaneous Localization and Mapping (SLAM) technologies. The aim is to establish a foundation for applications in knee arthroplasty. The emphasis is not on providing a complete markerless navigation solution, but on understanding the complexity and creating a groundwork upon which future approaches can be built.

A promising approach for markerless tracking is the use of photogrammetric techniques. Hu et al. (2021) demonstrated in a proof-of-concept study that scanning the knee joint and estimating its pose without markers is feasible. SLAM enables 3D reconstruction using cameras moving around the knee joint to capture all relevant surfaces. However, achieving high-precision reconstructions intraoperatively poses several challenges: the knee surface often lacks sufficient texture, reflections from wet tissue can degrade image quality, and the joint itself moves during surgery. To assess and optimize such methods, artificial bone models are often used because real medical datasets with ground truth are limited (Hu et al., 2021; He et al., 2022). While these models enable initial evaluations, they offer limited anatomical variability, and the reproducibility of experiments is constrained. Modifications to physical test setups often affect multiple parameters at once such as lighting, camera trajectory, and bone position complicating result interpretation. Moreover, realistic annotated images of bones are scarce, yet essential for AI-based segmentation and masking of the knee during SLAM (Neiss-Theuerkauff et al., 2024).

This work aims to develop and evaluate a trinocular SLAM system for intraoperative 3D reconstruction of the knee. To achieve this, a virtual camera-simulated environment is established under controlled conditions. Within this environment anatomical knee models are captured by a simulated trinocular camera system following defined motion.

The primary objective is to assess how camera configuration, motion, and viewpoint affect the results of the trinocular SLAM system, in detail the 3D reconstruction, camera motion and object localisation. As real surgical data with precise ground truth are not available, synthetic datasets are generated to provide reproducible reference data for quantitative evaluation. Furthermore, the work investigates whether object-specific segmentations of femur and tibia improve the accuracy of reconstructions.

While previous approaches have primarily focused on static knee models (Hu et al., 2021), this work extends SLAM evaluation to scenarios involving object motion. Using a trinocular camera system, we can acquire intraoperative

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-137-2026

movements of the knee joint and thus systematically test the reconstruction accuracy under dynamic conditions. This enables a controlled evaluation of SLAM performance in such situations.

2. Trinocular SLAM in Knee Arthroplasty

As an alternative to the classical navigation methods in knee arthroplasty, where markers are fixed in the bones and tracked with a stereo camera system, markerless approaches related to visual SLAM are being investigated. These approaches rely on 6-DoF pose tracking and 3D reconstruction. In this approach, a camera is moved around the knee joint and continuously captures image sequences, from which both the camera pose and a 3D reconstruction of the visible surfaces can be computed (Kahmen et al., 2020)

SLAM techniques originally stem from robotics and autonomous systems, where they are used for simultaneous mapping of unknown environments and self-localization. Well-known methods such as ORB-SLAM 2 (Mur-Artal and Tardos, 2017) were initially developed for mobile robots and augmented reality applications and are typically optimized for well-textured scenes in everyday environments. Direct transfer to the surgical context is challenging, as the algorithms must cope with specific difficulties such as smooth, reflective surfaces or object motion while meeting high demands for precision and real-time performance. These differences highlight the need for specially adapted SLAM solutions as well as systematic validation under the specific conditions of knee arthroplasty. To overcome these challenges, this study employs a trinocular camera setup, which provides enhanced perception and robustness in surgical scenarios (Conen et al., 2017).

2.1 Trinocular Camera System

Our system is designed for use in dynamic knee arthroplasty scenes. The foundation is a trinocular camera system, optimized based on the principles described by Kahmen et al. (2020). Three identical RGB cameras, the Basler ace acA1920-40uc, each equipped with a 25 mm Jade lens by Schneider Kreuznach, form the hardware framework for capturing the knee joint.

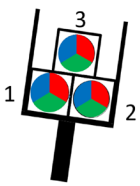

Figure 1. Schematic construction of a trinocular camera system (after Kahmen et al., 2020)

The cameras are arranged in an equilateral triangular layout (Figure 1.), with each camera slightly inclined toward the centre to maximize the overlapping field of view. This geometry improves depth accuracy and robustness of feature matching for dynamic scenes compared to conventional stereo setups (Maas, 1997; Conen et al., 2017). The inter-camera distance is approximately 10 cm, ensuring a compact and lightweight system suitable for use around the operating table.

All three cameras are hardware-synchronized to ensure simultaneous image capture across views. The system is pre-calibrated, providing both intrinsic parameters for each camera and the relative orientation between them. This calibration defines the fixed geometry of the trinocular camera system and enables consistent 3D reconstruction.

The optical design was selected considering the limited working distance in a sterile surgical field, which includes the disinfected operative site, sterile drapes, and the front sections of the sterile team members above table height (KRINKO, 2000). To avoid contamination of the sterile zone, the camera system is positioned at about 60 cm from the surgical site.

2.2 Trinocular SLAM

The trinocular SLAM workflow based on the work of COLMAP SLAM (Morelli et al., 2023) and Schierbaum et al. (2024). This workflow is designed for dynamic scenes in which multiple objects, in our application bones, and the calibrated camera system are in motion (Figure 2.). During each epoch, all three cameras simultaneously acquire synchronized image sequences of the scene. These images form the basis for 3D reconstruction and camera pose estimation.

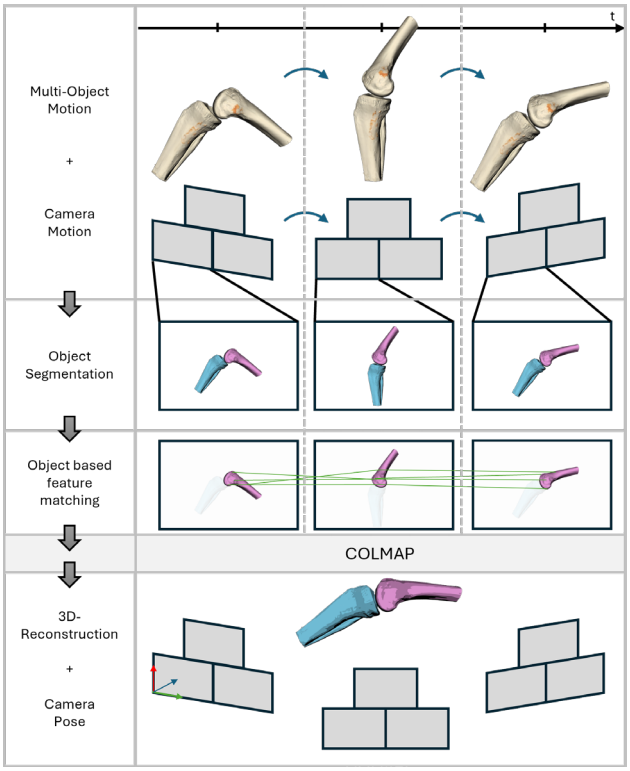

Figure 2. Workflow for trinocular SLAM with scene motion, and object segmentation: violet (master object), blue (slave object)

Segmenting the bones in the images is required to isolate individual objects for SLAM processing. While AI-based segmentation methods can be employed (Neiss-Theuerkauff et al., 2024), this work uses synthetic masks derived directly from the simulated environment. This approach guarantees consistent segmentation quality and minimizes errors that could arise from imperfect bone segmentation, allowing a focused evaluation of the SLAM pipeline.

Features are extracted for each segmented object using SuperPoint (DeTone et al., 2018), which provides robustness to viewpoint changes and image noise. Feature correspondences are established across images of the same epoch and with images from previous epochs using SuperGlue (Sarlin et al., 2020). By applying the segmentation masks, features are associated with their corresponding object, ensuring accurate object-specific matching.

The object-specific matches, together with the calibrated camera parameters, are imported into COLMAP, an open-source structure-from-motion (SfM) and multi-view stereo (MVS) framework (Schönberger and Frahm, 2016; Schönberger et al., 2016). COLMAP performs sequential reconstruction to estimate the camera pose relative to each object and generate dense point clouds. The scale of the reconstructed models is adjusted using the known relative orientation of the cameras.

To bring all objects into a common reference frame, the first object is defined as the master. The poses of all other objects (slaves) are then transformed relative to the master, using the inter-epoch camera motion (Figure 2.). Since the image triplets are synchronized, all camera poses can be expressed consistently in the first camera of the first epoch, resulting in a unified coordinate system for the entire scene.

This workflow allows the systematic evaluation of SLAM algorithms in dynamic scenarios, where multiple bones move independently. By combining synthetic masks, object-specific feature matching, and COLMAP-based reconstruction, the pipeline provides reproducible 3D reconstructions and camera poses for each object in a controlled, simulated surgical environment.

3. Creating Data in Virtual Environments

Due to the limited availability of real surgical data, generating representative test datasets remains a major challenge. Virtual environments provide a viable solution by enabling the creation of realistic and diverse datasets tailored to specific research objectives. In this work, the focus is on simulating the intraoperative optical acquisition of bone structures to provide reliable ground truth for evaluating SLAM algorithms and 3D reconstruction methods.

The software tool Blender (Blender Foundation, 2025) serves as the primary platform for this purpose. As open-source software for 3D modeling, animation, and visual effects, Blender combines an extensive graphical interface with the ability to automate complex tasks through Python scripting. This capability allows efficient generation of large and diverse datasets, which are essential for the systematic development and validation of optical imaging and navigation technologies. Using this approach, precise virtual bone models can be generated, ensuring reproducible and controlled scenarios for algorithm evaluation.

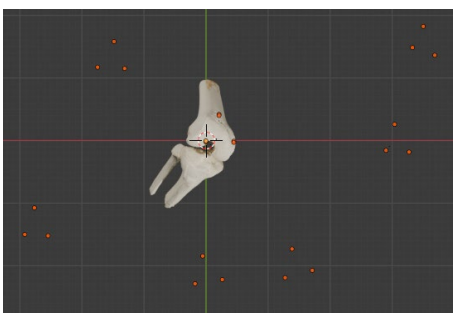

Figure 3. Viewpoints of the trinocular camera system on an anatomical 3D model in Blender. One cluster of three orange dots one position of the trinocular camera system.

To create the image sequences, each frame is rendered individually under settings that emulate realistic surgical conditions. In addition to technical adjustments related to camera, object, lighting, and rendering, the accurate positioning and alignment of the objects are crucial to ensure that the simulation meets the intended objectives. Figure 3. illustrates the setup of a scene with an anatomical knee model, demonstrating how these elements are integrated to form a realistic simulation environment.

3.1 Virtual Environment and Settings

3.1.1 Camera: For the generation of synthetic image data, the three Basler ace acA1920-40uc cameras were used as the basis of the multi-camera system. Each camera has a resolution of 1936 × 1216 pixels and a sensor size of 11.33 × 7.13 mm. The interior orientation, including principal point, principal distance, and lens distortion coefficients, was determined via photogrammetric self-calibration with relative orientation constraints in the software Ax.Ori (AXIOS 3D Services GmbH, 2010). Additionally, the relative orientation of the multi camera system was obtained during calibration. The camera parameters were transferred into Blender to replicate the real multi-camera system in the virtual environment. The focal length for the Blender camera model was calculated from the measured principal distance, sensor size, and image resolution. The principal point was adjusted to correctly represent the offset of the optical centre relative to the sensor. Lens distortions were not applied, as the Blender camera model does not natively support them.

The relative orientation of the multi-camera system was implemented by placing the first camera at the origin as a reference, while the remaining cameras were positioned according to the calibrated translations and deflections. In addition, the positions and orientations of the camera system were informed by a laboratory-based motion capture. Here, the actual movement of the trinocular camera system around a knee model was recorded, providing realistic trajectories and angular displacements. These motion data were then used to guide the placement and orientation of the cameras in the virtual environment, ensuring that the synthetic image sequences resemble real acquisition conditions. This approach ensures that the viewing directions of the real camera system are realistically reproduced in the virtual environment, providing a reliable basis for synthetic image generation.

3.1.2 Object: Physical artificial anatomical models for femur and tibia were scanned using structure-from-motion (SFM) techniques to create 3D reconstructions. These models were then aligned according to an anatomical knee joint, ensuring that the position and orientation of the bones correspond to real anatomy (Figure 4.). To provide additional contextualization, an open skin leg model was created, serving as a reference for correct placement of the bones and to limit the visible area.

Figure 4. Artificial knee model placed within a modelled leg

All other anatomical structures are omitted, resulting in images with only the knee is visible, simulating the surgery conditions. This approach allows the image data to be used specifically for the evaluation of SLAM algorithms, without interfering elements or additional structures affecting the reconstruction.

3.1.3 Lighting: The illumination of the surgical field is subject to specific requirements defined by the IEC 60601-2-41 standard (International Electrotechnical Commission, 2023). Modern surgical lights with these requirements and are approved for medical use. Consequently, these characteristics should replicate within the virtual Blender environment.

In Blender, light intensity is specified in watts by default (Blender Foundation, 2025), which complicates a direct implementation of the specifications of real lights. Due to this limitation, the adjustment of the virtual lighting is primarily based on visual assessment. This approach inevitably deviates from both the exact properties of the real surgical light and the requirements defined in the standard. Within the virtual environment, the lighting is simulated using two spotlights. They are positioned at 1330 mm from the object and oriented to provide uniform and intensive illumination of the surgical field. With the camera, objects, and lighting accurately defined, the next step is to render the synthetic images.

3.1.4 Rendering: For rendering the synthetic datasets, Blender’s Cycles engine is used, as it relies on physically ray-traced calculations and thus enables particularly realistic light and material effects. To ensure consistent image quality, the number of samples was set to 8. Denoising and adaptive sampling were deliberately omitted to preserve the raw data as unaltered as possible (Blender Foundation, 2025). The images are output in PNG format with an 8-bit colour depth, providing lossless storage while keeping file sizes manageable. These settings are also aligned with the parameters of real cameras and providing a controlled basis for reliably comparing the synthetic image data with real recordings. Object-specific segmentation masks were also generated using Vision Blender (Cartucho et al. 2021), assigning unique identifiers to each bone and producing pixel-accurate masks directly from the 3D scene, thus supplying ground truth for SLAM evaluation.

3.2 Dataset

The dataset replicates the acquisition of the knee joint using the trinocular camera system. For this image triplets are generated, considering the camera settings derived from the photogrammetric calibration. In the rendered images the visible regions of the knee surface are included. The surrounding skin or areas occluded by it are not displayed, while the knee joint itself remains static during acquisition, and only the camera system moves. In addition, we generate for each image two masks, one for the femur and one for the tibia. This effectively isolating the relevant structures for SLAM evaluation.

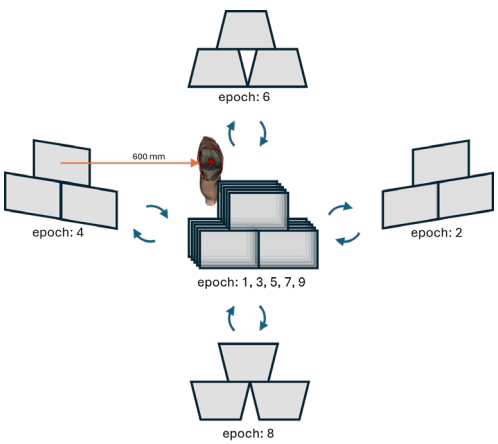

Figure 5. Synthetic image acquisition with varying deflection in a cross-shaped arrangement

The synthetic dataset is generated to assess the performance of SLAM algorithms under controlled and reproducible conditions. Their purpose is to systematically investigate how different camera motion and viewpoints affect the coverage and quality of the reconstructed knee joint. The goal is to identify acquisition configurations that provide sufficient overlap and completeness while minimizing the number of required images.

For this purpose, the camera system moves at a constant distance of 60cm around the object, following a cross-shaped pattern (Figure 5.) that extends from a central starting position right, left, upward and downward. The dataset is based on the angular difference between the central starting position and extreme orientations. For each direction, images are captured at the maximum angles (+ and –) as well as at the central position, resulting in five defined camera positions and nine image triplets in total. Varying the deflection angles between 5° and 45° defines the spacing between successive image triplets (Table 1.), allowing conclusions to be drawn about suitable real-world acquisition trajectories.

No.	Deflection [°]	Distance between two epochs [mm]
1	5	52.3
2	15	156.6
3	25	259.7
4	35	360.8
5	45	459.2

Table 1. Deflection and the corresponding camera motion

4. Results of Trinocular SLAM

4.1 Camera Pose

For the evaluation of the dataset, the camera poses were first analysed by computing the Euclidean distance between the estimated camera positions and the ground truth. Since the knee joint remains static in this dataset, the camera poses can be determined for the entire knee, the femur, and the tibia (Table 2.). The same ground truth is used for all evaluations. The following table lists the mean deviations as well as the maximum deviations from the known position for all epochs with one deflection angle, for each segmented structure.

No.	Object	Deflection [°]	RMS 3D [mm]	Max 3D [mm]
1-a	knee	5	0.60	1.35
1-b	femur	5	0.85	1.84
1-c	tibia	5	0.77	1.88
2-a	knee	15	0.80	2.07
2-b	femur	15	1.19	2.84
2-c	tibia	15	0.85	2.17
3-a	knee	25	1.25	3.39
3-b	femur	25	1.78	4.89
3-c	tibia	25	1.28	3.88
4-a*	knee	35	2.31	5.58
4-b*	femur	35	2.79	8.20
4-c*	tibia	35	2.50	7.13
5-a*	knee	45	2.68	6.96
5-b*	femur	45	3.74	9.76
5-c*	tibia	45	3.32	7.40

Table 2. Camera poses deviation in relation to ground truth. Datasets marked with * camera poses excluding the two last epochs

https://doi.org/10.5194/isprs-annals-XI-2-2026-137-2026

The lowest mean deviations from the reference were observed for the sequences with 5° deflection around the artificial knee joint. For the pose determination based on the entire knee, the mean deviation is 0.60 mm (No. 1-a). The deviations for the femur and tibia evaluations are 0.85 mm (1-b) and 0.77 mm (1-c), respectively, also remaining below one millimetre on average.

Larger deflections of the camera system around the knee lead to higher mean deviations from the reference. The trend remains that deviations are smaller when computed over the entire knee joint, followed consistently by the tibia. For the evaluations at 35° and 45° deflection, only seven instead of nine positions were included, as the evaluation was stopped after the 7th epoch in these cases. This is since the femur is not visible in the data from the top view (see Figure 9.) and these images were not filtered out prior to processing.

4.2 Object Pose

As shown in the trinocular SLAM workflow, the oriented camera poses serve as the basis for generating the dense point clouds. These point clouds are represented in the same coordinate system as the camera poses. Since the viewpoints and orientations of both the camera system and the synthetic knee joint are known in the virtual environment, the object poses relative to the ground truth can be determined. The deviations in 6DOF (Table 3.) were computed using the cloud-to-cloud based ICP algorithm implemented in CloudCompare (CloudCompare, 2022).

No.	Deflection [°]	No. of Image Triple	Knee (a)	Femur (b)	Tibia (c)
			3D [mm]	3D [mm]	3D [mm]
1.2	5	2	6.60	6.31	18.11
1.3	5	3	6.38	5.48	15.57
1.4	5	4	6.23	5.71	16.20
1.5	5	5	5.93	7.23	15.08
2.2	15	2	5.07	3.81	**1.55**
2.3	15	3	4.14	2.08	1.94
2.4	15	4	3.79	**1.84**	3.90
2.5	15	5	3.29	2.21	4.36
3.2	25	2	6.89	4.48	2.50
3.3	25	3	3.23	3.16	2.44
3.4	25	4	3.26	2.78	3.77
3.5	25	5	**2.29**	2.53	3.73
4.2	35	2	6.17	5.31	3.80
4.3	35	3	3.05	4.04	1.99
4.4	35	4	3.43	3.82	3.81
4.5	35	5		interrupt	
5.2	45	2	7.73	10.23	8.13
5.3	45	3	5.05	8.45	5.51
5.4	45	4	4.11	6.64	6.33
5.5	45	5		interrupt	

Table 3. Absolute 3D deviations of the object position to the ground truth

The results indicate substantial differences in global deviations depending on the segmentation used. Even within the same object, deviations vary considerably. For the evaluation based on the entire knee joint, the minimum deviation of 2.29 mm occurs at 25° deflection with five image triplets (No. 3.5-a). The maximum 3D deviation is 7.73 mm (No. 5.2-a). For the femur, deviations range from 1.84 mm (No. 2.4-b) to 10.23 mm (No. 5.2-a), while the tibia exhibits the largest overall range, from 1.55 mm (No. 2.2-c) to 18.11 mm (No. 1.1-c). Notably, the 5° deflection sequences for the tibia show particularly high deviations (No. 1.2-c – 1.5-c). Similarly, the 5° deflection leads to higher deviations for the femur and the entire knee joint.

When considering all deflections together, the 15° deflection yields the lowest deviations, followed by the 25° deflection. Evaluations at 35° and 45° deflection with only five image triplets could not be completed. In general, deviations are highest when only two image triplets are used and decrease with additional triplets, except for the tibia, which does not follow this trend.

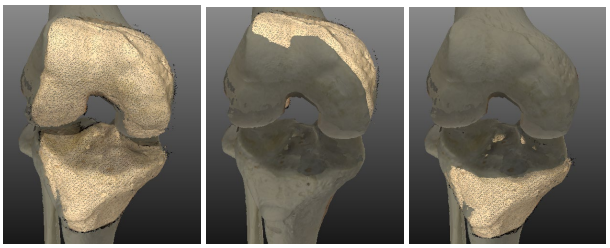

Figure 6. Offset representation of the resulting point clouds (highlighted) of knee (left), femur (centre) and tibia (right) relative to ground truth (No. 3.3).

Figure 6. provides an example of the resulting point clouds for the knee joint, femur, and tibia, along with their offsets relative to the ground truth. It is noticeable that the deviations manifest differently along the coordinate axes, with the largest discrepancies occurring along the Z-axis, i.e., the viewing direction of the camera system.

4.3 Object Reconstruction

By applying the ICP-based alignment between the reconstructed point clouds and the reference surface model, not only absolute positional deviations can be quantified, but also the overall reconstruction quality of the SLAM results can be analysed. This analysis provides spatially resolved information about local reconstruction errors, offering additional insights beyond global deviation metrics.

Using CloudCompare (CloudCompare, 2022), the cloud-to-mesh distance was computed to visualize local deviations from the reference model, which served as the input geometry during the synthetic data generation. The notation (x.y) refers to the respective test configuration as listed in Table 3., where “y” indicates the number of image triplets used.

Distinct differences are observed between the knee (a) and segmented (b, c) evaluations (Figure 7.). In the knee reconstructions, local deviations of up to approximately 0.5 mm occur on both the femur and tibia surfaces, affecting larger contiguous regions. In contrast, the segment-based reconstructions show deviations primarily in border areas or local surface extrema, particularly on the tibia, while the femur remains comparatively stable. Only marginal improvements are visible beyond the third image triplet (1.3), with the most notable deviation appearing in the two-triplet configuration (1.2), where the upper femur region appears thinner reconstructed compared to the reference model.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-137-2026 | © Author(s) 2026. CC BY 4.0 License.

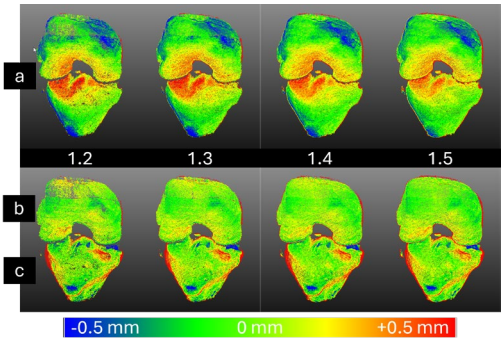

Figure 7. Coloured results of the 5° camera motion. The upper row (a) shows reconstructions of the entire knee based on two (1.2) to five (1.5) image triplets. The lower rows (b) and (c) depict the corresponding reconstructions of the femur and tibia.

The results for the 25° camera deflection (Figure 8.) differ noticeably from those observed for the 5° deflection. When evaluating the full knee, the overall deviations are smaller, while the affected surface regions remain largely consistent. The local deviations observed in the reconstructions decrease with the inclusion of additional image triplets. For the separate reconstructions of the femur and tibia at this camera motion, only minor deviations remain. Again, a consistent trend can be observed: the reconstruction accuracy improves as more image triplets are incorporated.

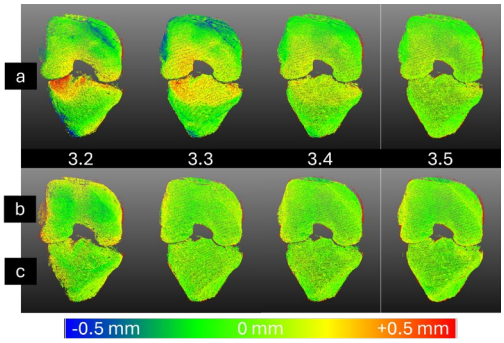

Figure 8. Coloured results of the 25° camera deflection. The upper row (a) shows reconstructions of the entire knee based on two (3.2) to five (3.5) image triplets. The lower rows (b) and (c) depict the corresponding reconstructions of the femur and tibia.

In the additional datasets with camera deflection s of 15°, 35°, and 45° (see Appendix), comparable tendencies can be identified. The reconstruction accuracy benefits from the inclusion of at least three image triplets. Furthermore, deviations are generally smaller for individually reconstructed bones. Observations from the 35° and 45° datasets show that certain surface regions cannot be reconstructed when only a few image triplets are available, and no complete reconstruction could be obtained using five image triplets.

5. Discussion

A comprehensive analysis of the results for camera pose, object pose, and object reconstruction allows conclusions to be drawn regarding the performance and quality of the trinocular SLAM approach based on the synthetic dataset. The accuracy of the estimated camera poses is closely related to the camera motion. With larger deflections between consecutive image triplets, the overall deviations increase. Estimating the system orientation using the entire knee provides an advantage in accuracy, as a larger image area is available for feature matching. Similar tendencies have been reported in previous studies (Schierbaum et al., 2024).

With relation to the individual bones, would expect the femur to yield more accurate pose estimations than the tibia, given its larger surface. However, this is not the case, which may be attributed to the limited visibility of the femur in the images captured by the third (upper) camera. As illustrated in Figure 9., the femur is not always fully captured in the field of view for certain deflections.

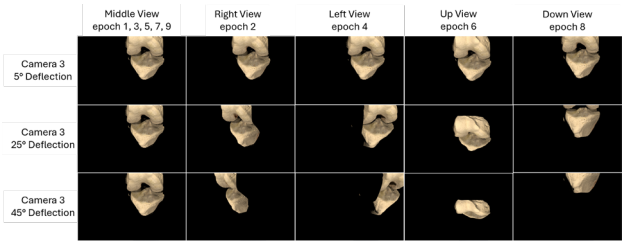

Figure 9. Synthetic views from the third (upper) camera of the trinocular system at deflection of 5°, 25°, and 45° around the knee.

The quality of the camera pose cannot be directly transferred to the reconstruction quality. While the 5° deflection yields the most accurate camera poses, the corresponding results in terms of surface and object alignment are inferior to those of the 15° and 25° datasets.

To evaluate the object pose and object reconstruction, it is useful to consider both together, since they are derived using the same ICP-based procedure. It was expected that smaller camera motions would lead to larger deviations, as positional offsets, particularly along the viewing axis, become more pronounced. In this direction, the uncertainty of the reconstructed surface is also higher, as it is more sensitive to the underlying image measurement accuracy.

Moreover, deviations in the object pose correlate with the surface reconstruction quality, as the ICP algorithm determines the transformation between the reconstructed surface and the ground truth model based on geometric alignment. Considering this relationship, one can conclude that even though additional image triplets reduce local surface deviations, they do not necessarily improve the estimated object pose.

Overall, both surface accuracy and camera pose precision influence the resulting spatial offset of the reconstructed object. Within the analysed datasets, the 15° and 25° camera deflections provide the most consistent and accurate results, whereas the other configurations yield lower quality in different aspects.

When comparing the evaluation based on separate masking of the femur and tibia with that using a combined mask of the entire knee, it becomes evident that the surfaces are reconstructed more accurately when masking individually. The positional deviations are also slightly smaller for the 15° and 25° datasets in the segmented evaluation compared to the full-knee reconstruction.

Regardless of the masking strategy, it can be observed that reconstructions show smaller deviations when a larger number of image triplets are included in the point cloud generation.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-137-2026 | © Author(s) 2026. CC BY 4.0 License.

However, the benefit of additional triplets reaches a limit rather quickly, as three triplets often provide a reconstruction quality comparable to that obtained with five.

A particularly important aspect in terms of local reconstruction accuracy is the coverage of peripheral regions, since deviations tend to accumulate at the edges of the reconstructed surfaces. Alternatively, these border areas could be filtered or excluded from the analysis to minimize their influence. In this context, it also remains an open question how large the reconstructed surface area must be to achieve a sufficiently precise registration of the bones. This is an aspect that cannot be conclusively derived from the present results.

Overall, the dataset provides valuable insights into the performance of trinocular SLAM for knee joint acquisition. It offers information on the optimal camera motion between two image triplets required for accurate pose estimation, as well as the number of triplets needed to achieve a sufficiently detailed surface reconstruction. Moreover, the synthetic data allow the generation of precise ground truth information, which serves as a reliable basis for evaluating reconstruction results.

However, several limitations must be considered. The findings cannot be directly transferred to real surgical data. The underlying 3D model of the artificial knee exhibits homogeneous surface textures, which may simplify feature matching and thus overestimate SLAM robustness. In addition, external influences such as lighting variations, reflections, and sensor noise are not represented in the virtual environment. The synthetic images also lack optical distortions present in real cameras, which can affect calibration and matching performance. Moreover, the intended application of trinocular SLAM was only partially addressed in this study, as the bones within the dataset remained static, the influence of object motion could therefore not be evaluated.

In the broader scientific context, this approach extends existing research on vision-based SLAM systems in medical applications. Previous works have primarily evaluated SLAM on static anatomical models, where reproducibility is limited and ground-truth data are often unavailable (Hu et al., 2021; Kahmen et al., 2020; Schierbaum et al., 2024). While such works provide valuable feasibility insights, they do not allow systematic investigation of how camera motion, segmentation, and object movement affect reconstruction accuracy.

The virtual simulation environment introduced in this work addresses these gaps by enabling reproducible and parameter-controlled testing of a trinocular multi-object SLAM system. Unlike conventional single-object SLAM approaches, the proposed framework explicitly accounts for multiple independently moving structures and thus better reflects the complexity of surgical scenes. This study therefore establishes an important intermediate step between algorithmic research in robotics and clinical applications of SLAM-based navigation systems.

6. Conclusion and Future Work

This work set out to develop and evaluate a trinocular SLAM framework for intraoperative 3D reconstruction of the knee using a controlled virtual simulation environment. Within this framework, synthetic datasets were created to enable systematic analysis under defined conditions and to provide reproducible ground-truth data that are not available in real surgical scenarios.

The objectives defined in this work were successfully achieved. The evaluation demonstrated that both camera motion and configuration have a substantial influence on SLAM performance. Specifically, moderate rotations of 15° to 25° combined with at least three image triplets produced the most accurate camera poses and surface reconstructions. Furthermore, object-specific segmentation of the femur and tibia proved to enhance reconstruction accuracy by reducing local deviations compared to reconstructions based on the full-knee mask. These findings validate the use of synthetic data for controlled benchmarking and provide valuable insights into how acquisition strategies affect reconstruction outcomes. While the synthetic setup enables precise, repeatable testing, it also introduces limitations in realism. Nevertheless, the results establish an essential foundation for future research toward markerless navigation and intraoperative 3D surface tracking in knee arthroplasty.

Future work will focus on extending the dataset to include dynamic knee motion and increasing realism through 3D Gaussian Splatting. Experiments with real image sequences will follow, using an external tracking system to provide ground-truth trajectories. AI-based segmentation of bone structures will replace predefined masks, enabling automated and scalable processing.

Ultimately, the trinocular SLAM system is to be adapted for application in total knee arthroplasty. Beyond accurate reconstruction, it should enable real-time tracking of bone motion and the integration of preoperative planning into the surgical workflow, paving the way toward fully markerless, computer-assisted orthopedic surgery.

Acknowledgments

This work was funded by the Federal Ministry of Education and Research (BMBF). We would like to thank Aesculap and AXIOS 3D Services for their co-funding, and the PIUS Hospital, Oldenburg, for their insights and expertise.

References

AXIOS 3D Services GmbH, 2010: Ax.Ori, Version 1.10. axios3d.de (12 June 2025).

Blender Foundation, 2025: Blender, Version 4.4. blender.org (30 October 2025).

Cartucho, J., Tukra, S., Li, Y., Elson, D. S., Giannarou, S., 2021: VisionBlender: a tool to efficiently generate computer vision datasets for robotic surgery. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 9(4), 331–338. doi:10.1080/21681163.2020.1835546.

CloudCompare, 2022: CloudCompare, Version 2.12.4. cloudcompare.org (30 October 2025).

Conen, N., Luhmann, T., and Maas, H.-G., 2017: Development and Evaluation of a Miniature Trinocular Camera System for Surgical Measurement Applications. PFG, 85(2), 127–138. doi:10.1007/s41064-017-0014-3.

DeTone, D., Malisiewicz, T., Rabinovich, A., 2017: SuperPoint: Self-Supervised Interest Point Detection and Description. arXiv. doi:10.48550/ARXIV.1712.07629.

He, G., Ricca, J. M., Dai, A. Z., Mustahsan, V. M., Cai, Y., Bielski, M. R., Kao, I., Khan, F. A., 2022: A novel bone registration method using impression molding and structured-light 3D scanning technology. Journal of Orthopaedic Research, 40, 2340–2349. doi:10.1002/jor.25275

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:10.1109/ACCESS.2021.3075628.

International Electrotechnical Commission, 2023: Medical electrical equipment – Part 2-41: Particular requirements for the basic safety and essential performance of surgical luminaires and luminaires for diagnosis (IEC 60601-2-41:2021).

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.*, XLIII-B2-2020, 35–42. doi:10.5194/isprs-archives-XLIII-B2-2020-35-2020.

KRINKO, 2000: Anforderungen der Hygiene bei Operationen und anderen invasiven Eingriffen. Robert Koch-Institut, Apr. 2000. doi:10.25646/162.

Maas, H.-G., 1997: Mehrbildtechniken in der digitalen Photogrammetrie, 118 p. doi:10.3929/ETHZ-A-001865074.

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., XLVIII-1/W1-2023, 317–324. doi:10.5194/isprs-archives-XLVIII-1-W1-2023-317-2023.

Moret, C. S., Hirschmann, M. T., 2021: Navigation und Robotik in der Knieendoprothetik. Arthroskopie, 34(5), 351–357. doi:10.1007/s00142-021-00467-6.

Mur-Artal, R., Tardos, J. D., 2017: ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. IEEE Trans. Robot., 33(5), 1255–1262. doi:10.1109/TRO.2017.2705103.

Neiss-Theuerkauff, T., Schierbaum, A., Luhmann, T., Sieberth, T., Wallhoff, F., 2024: Semantic Bone Structure Segmentation in 2D Image Data: Towards Total Knee Arthroplasty. In: Bramer, M., Stahl, F., eds. Artificial Intelligence XLI, Lecture Notes in Computer Science, 15446, 352–357. Cham: Springer Nature Switzerland. doi:10.1007/978-3-031-77915-2_29.

Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020: SuperGlue: Learning Feature Matching With Graph Neural Networks. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 4937–4949. doi:10.1109/CVPR42600.2020.00499.

Schierbaum, A., Neiss-Theuerkauff, T., Luhmann, T., Wallhoff, F., Sieberth, T., 2024: Investigations on 3D reconstruction of bones in surgery using a handheld trinocular camera system. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII-2/W7-2024, 145–151. doi:10.5194/isprs-archives-XLVIII-2-W7-2024-145-2024.

Schönberger, J. L., Zheng, E., Frahm, J.-M., Pollefeys, M., 2016: Pixelwise View Selection for Unstructured Multi-View Stereo. In: Leibe, B., Matas, J., Sebe, N., Welling, M., eds. Computer Vision – ECCV 2016, Lecture Notes in Computer Science, 9907, 501–518. Cham: Springer International Publishing. doi:10.1007/978-3-319-46487-9_31.

Schönberger, J. L., Frahm, J.-M., 2016: Structure-from-Motion Revisited. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 4104–4113. doi:10.1109/CVPR.2016.445.

Stübig, T., Windhagen, H., Krettek, C., Ettinger, M., 2020: Computer-Assisted Orthopedic and Trauma Surgery. Deutsches Ärzteblatt International, Nov. 2020. doi:10.3238/arztebl.2020.0793.

Appendix

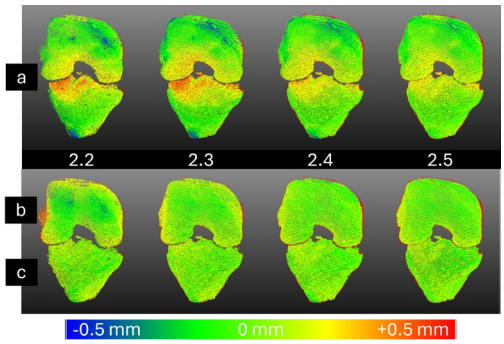

Figure 10. Coloured results of the 15° camera motion. The upper row (a) shows reconstructions of the entire knee based on two (2.2) to five (2.5) image triplets. The lower rows (b) and (c) depict the corresponding reconstructions of the femur and tibia.

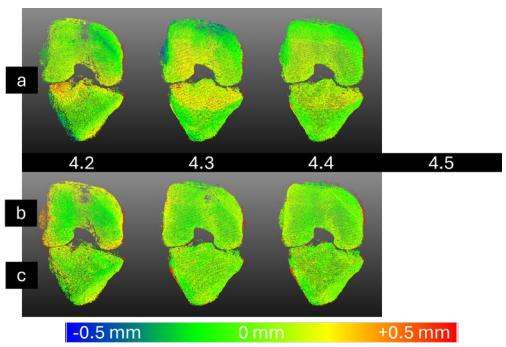

Figure 11. Coloured results of the 35° camera motion.

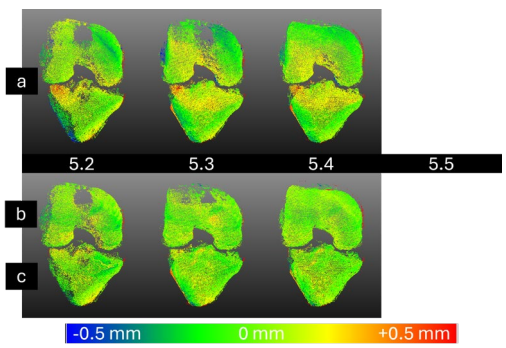

Figure 12. Coloured results of the 45° camera motion.