Accuracy and Reliability of an Augmented Reality Prototype for Robot-Assisted Partial Nephrectomy: A Preclinical Study with 3D-printed Kidney Phantoms (UroCCR 168)

F. Rubat Baleuri*[1,2], M. Pattou[1,2], G. Margue[1,2,3], G. Nguyen[4], A. Pitout[1,2], A. Khaddad[1,2], K. Chandelon[4], J. Desternes[4], N. Bourdel[4,5], J.-C. Bernhard[1,2,3], A. Bartoli[4,5,6]

*: Corresponding author. Department of Urology, University Hospital, Bordeaux, France

Affiliations:

1: Urology Department, Bordeaux University Hospital, 33000 Bordeaux, France
2: I.CaRe Bordeaux—BRIC Inserm U1312, 33200 Bordeaux, France
3: French AFU Cancer Committee Guidelines, 75017 Paris, France
4: SurgAR, 63000 Clermont-Ferrand, France
5: EnCoV, UMR 6602 Institut Pascal, CNRS – University Clermont Auvergne, 63000 Clermont-Ferrand, France
6: Department of Clinical Research and Innovation, Clermont-Ferrand University Hospital, 63000 Clermont-Ferrand, France

*Corresponding author: Federico RUBAT BALEURI (F.R.B.) - federico.rubat@gmail.com

ORCID: 0009-0006-9724-8353

Funding: This work has benefited from a government grant managed by the French national research agency (ANR) under the third investment program for the future (PIA), part of France 2030, reference ANR-21-RHUS-0015 and for a European grant managed by the Region Nouvelle Aquitaine.

ABSTRACT

Introduction

Augmented Reality (AR) systems can enhance intraoperative anatomical understanding during Robot-Assisted Partial Nephrectomy by overlaying preoperative 3D models onto the surgical view. Their usefulness, however, depends on the spatial accuracy of these overlays, which requires objective validation.

Methods

This preclinical study quantitatively evaluated the spatial accuracy of a new AR software prototype using a validation workflow based on 3D-printed kidney phantoms and a Structure-from-Motion (SfM) reconstruction as ground truth. CT scans were segmented to create virtual kidney models that were 3D printed and filmed with the da Vinci Si camera. The AR prototype was applied to the recordings and manual 3D registration was performed using the TilePro interface. Keyframes from manual and tracking-based registrations were analysed to compute the Target Registration Error (TRE), defined as the mean Euclidean distance between corresponding vertices of the AR and SfM models. Three operators repeated the workflow and manual registration time was recorded to assess variability and feasibility.

Results

Mean TRE across experiments was below 5 mm, indicating clinically acceptable accuracy. TRE values for individual keyframes remained consistently under 10 mm, with limited inter-operator variability (ICC(3,*k*) = 0.80). Manual and tracking TREs were comparable and strongly correlated (β = 0.89; $p < 0.001$). Tracking-related precision changes varied among operators and mean manual registration time was 173.6 ± 76.4 s. Accuracy was unaffected by keyframe count, kidney variability, or registration duration.

Conclusion

The proposed methodology offers an objective, reproducible framework for assessing AR registration in preclinical settings. The prototype showed reliable spatial performance, supporting progression toward in vivo validation.

Keywords: Augmented Reality; Robot-Assisted Partial Nephrectomy; 3D Registration; Structure-from-Motion; Target Registration Error; Surgical Navigation; Validation Methodology

INTRODUCTION

Renal Cell Carcinoma (RCC) is a common urologic malignancy, with a rising incidence due in part to improved imaging and earlier detection [1,2]. When technically feasible, Robot-Assisted Partial Nephrectomy (RAPN) is the preferred surgical approach for localized RCC, offering favourable perioperative and functional outcomes compared to open surgery [3–5]. However, RAPN remains technically demanding, particularly for tumours with complex anatomical relationships to vascular and urinary structures [6]. Achieving optimal surgical outcomes, commonly defined by the trifecta of negative margins, short warm ischemia time and absence of major complications, requires precise preoperative planning and accurate intraoperative anatomical guidance [7–9].

To address the cognitive burden of interpreting traditional two-dimensional imaging, Virtual Reality (VR) 3D reconstructions from preoperative CT or MRI data have gained traction in recent years [10–12]. Further, these VR reconstructions can now be integrated into Augmented Reality (AR) systems, enhancing intraoperative navigation by superimposing patient-specific virtual anatomy onto the surgical field [13]. Multiple clinical studies suggest that image-guided technologies may improve surgical performance and outcomes in partial nephrectomy [14].

Therefore, AR systems require validation of the spatial accuracy with which they display the digital content to the surgeon, as this directly determines the reliability of their use for intraoperative guidance, particularly in the demanding context of oncologic surgery. Target Registration Error (TRE) is a widely accepted metric for assessing the accuracy of AR systems that is defined as the Euclidean distance between corresponding target points in the virtual reconstruction and in the organ it derives from, after registration of the two [15–17]. Several methodologies exist to estimate TRE, including analytical models [18], empirical measurements using physical phantoms [19], simulation-based approaches [20], and uncertainty modelling [21].

This preclinical study primarily aimed to quantitatively evaluate the spatial accuracy of a novel AR prototype developed for RAPN in a controlled experimental setting, in order to contextualise future evaluations in live surgical environments. Secondarily, it aimed to evaluate the feasibility and repeatability of the newly developed evaluation methodology, also for its possible future use in live surgery.

METHODOLOGY

All patients in this study provided informed consent and were subsequently enrolled in the UroCCR database, a comprehensive national multicentre registry that prospectively captures imaging, surgical videos, clinical, biological and radiological data on kidney cancer (ClinicalTrials.gov: NCT03293563/CNIL agreement DR-2013-206).

The experimental methodology in this study is based on empirical measurements of the TRE in a controlled preclinical setting. The approach relies on the Structure-from-Motion (SfM) reconstruction technique, a photogrammetric method that enables intraoperative reconstruction of a 3D surface model of the organ from images or videos recorded with the surgical camera. Two types of images were used for SfM reconstruction. The first type consists of *keyframes*, that are images acquired with the K-SurgAR AR prototype during the initial manual registration and subsequent tracking phases. These keyframes will serve as reference points for automatically registering the virtual model onto the 3D-printed model and are defined as a frame in which the spatial relationship between the

endoscope pose, meaning its position and orientation, and the target anatomy is known with high accuracy. The second type consists of *enrichment frames*, that are images extracted from the surgical video independently of the AR prototype. Their purpose is to increase the number of images available during the photogrammetric process, thereby improving the accuracy of the resulting reconstruction.

SfM-based 3D reconstructions of ten anatomically accurate 3D-printed kidney phantoms were used as ground truth and compared with the corresponding virtual kidney models displayed by the AR prototype after manual and automatic tracking-based registration. To ensure clinical relevance, TRE calculations focused on the tumour–parenchyma junction at the kidney surface, a region that is critical during the enucleation phase of partial nephrectomy and therefore represents the area where registration of the virtual model must be most precise.

As the primary objective for spatial accuracy evaluation, an acceptability threshold of 5 mm was established for the TRE. This threshold was considered appropriate for the proposed application, given the typical surgical margin size adopted in partial nephrectomies across different centres and our newly proposed methodology. Previous works about image guided surgery found mean TRE between 2.2 (±1.3) mm and 3.8 (±3.7) mm accordingly to a review, depending on the methodology and the surgical speciality [22], while a previous work specifically concerning 3D-IGRAPN found a TRE of 4.88mm for exophytic tumours [23]. Additional evaluation criteria included the mean TRE, used for the analysis of tracking results, in which a TRE value was available for each keyframe. The Delta TRE (manual – tracking) was instead used to quantify the potential precision gain of the tracking phase compared to the corresponding manual phase.

The feasibility of the proposed workflow was evaluated by recording manual registration times; additionally, the association between the time required for the registration process and the resulting TRE was assessed. The reliability of the workflow was defined as the inter-operator reproducibility of TRE computation, quantified using the Intraclass Correlation Coefficient (ICC) and interpreted according to the thresholds proposed by Koo & Li (2016). ICC values ≥ 0.75 were considered indicative of an acceptable level of reliability for experimental clinical applications. Repeatability was further assessed by having three independent operators replicate the process and analysing inter-operator variability. An overview of the workflow is shown in Figure 1, with further details provided in the following sections.

Descriptive statistics were used to summarize TRE values and registration times and are reported as mean ± standard deviation. Inter-operator reproducibility of manual TRE measurements was assessed using a two-way mixed-effects intraclass correlation coefficient for absolute agreement (ICC[3,*k*]), with 95% confidence intervals, and interpreted according to the criteria proposed by Koo and Li. Associations between continuous variables, including manual TRE, tracking TRE, number of tracked keyframes, and registration time, were evaluated using Pearson's correlation coefficient. To account for repeated measurements obtained from the same kidney phantom during tracking, a linear mixed-effects model was applied, including the kidney phantom as a random effect. All statistical tests were two-sided, and a p value < 0.05 was considered statistically significant. Statistical analyses were performed using R studio statistical software.

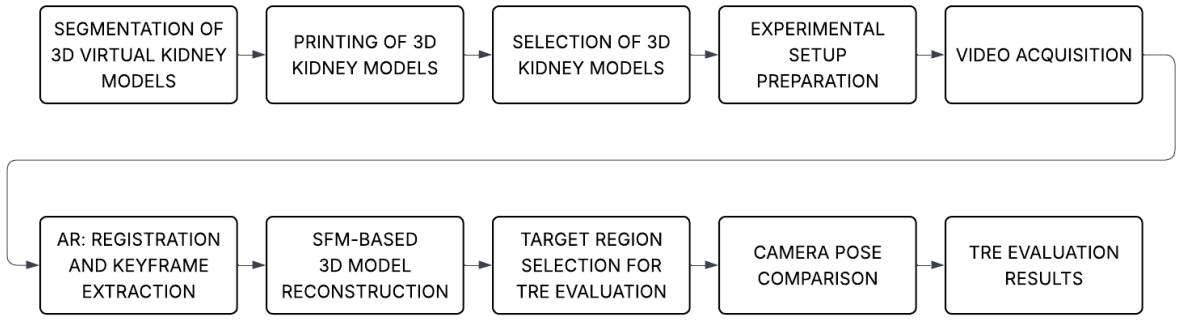

Figure 1: Methodology Flowchart.

1. Kidney Phantoms

Patient-specific kidney models were created from preoperative multiphase CT scans using Synapse 3D® Kidney Analysis software (Fujifilm, Tokyo, Japan).

The CT-scan used for the 3D model reconstruction was usually the one made for kidney tumour diagnosis. If the primary CT scan did not provide an angiogram, a urinary phase, or if the slices were thicker than 1.5 mm, a secondary CT scan was made at our institution with a specific protocol: 4 phase MDC scanning, including arterial, cortico-medullary, and excretory phases, with 0.6 mm slices.

Segmentations included the arterial and venous trees, tumour volume, collecting system, and renal parenchyma. The STL files of the kidney models were then exported and used to produce transparent, rigid 3D-printed models via a Stratasys J750 multi-material printer. The process has already been described and validated in previous works [24,25].

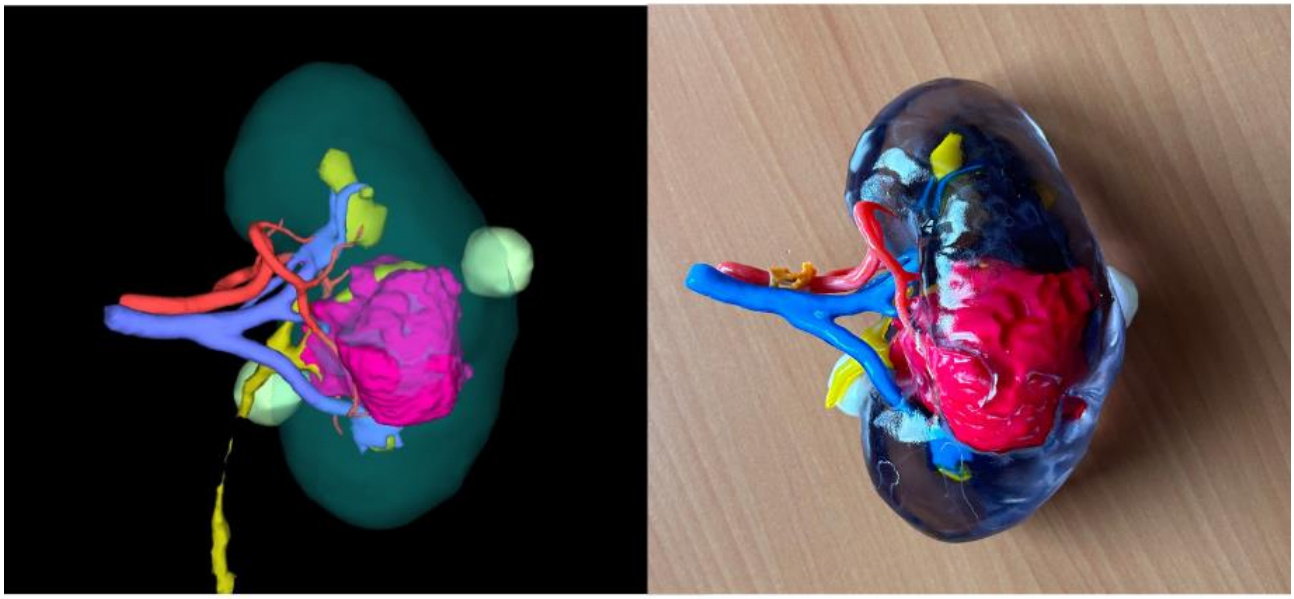

Figure 2: Example of virtual kidney model on the left and corresponding 3D-printed kidney model on the right.

Ten kidney models were selected to represent a range of anatomical variations and tumour complexities, including differences in size, location, laterality, and growth patterns. The transparent

rigid prints were coated with AESUB spray to minimise surface reflections and enhance 3D tracking performance. A detailed description of the selected models is shown in Table 1.

Kidney ID	Tumour Size (cm)	Kidney	Location	Sinus Contact	Growth Pattern	RENAL score
3	8	Right	Upper pole	Direct contact	<50% endophytic	10xh
7	3,5	Right	Lower pole	No	>50% endophytic	5a
9	5,5	Left	Equatorial	Small contact	>50% endophytic	9a
11	1,7 + 1,6 + 0,9	Right	Lower pole	No	Predominantly exophytic	4p
13	2,8	Left	Upper pole	No	Predominantly endophytic	8x
42	8,1	Right	Upper pole	No	Predominantly exophytic	6x
5009	3,5	Right	Equatorial	No	>50% endophytic	5p
5021	4,5	Right	Equatorial	No	<50% endophytic	8a
5026	5,5	Left	Lower pole	Direct contact	Predominantly exophytic	7a
5065	4,5	Left	Equatorial	Direct contact	Predominantly endophytic	10a

Table 1: Description of the 3D-printed kidney models.

2. Video Acquisition and Keyframe Registration

The experimental setup used to record the video footage was equipped with a robot-assisted surgery simulation platform, combining a Da Vinci Si robotic system, a laparoscopic Pelvitrainer and a Medicapture stereoscopic video acquisition system. Kidney phantoms were placed inside the Pelvitrainer showing the anterior face, the posterior face, or the convexity, depending on which configuration offered the clearest view of the tumour. An ArUco marker [26] was placed adjacent to each model to provide a fixed reference for spatial scaling and to aid SfM reconstruction.

Prior to video acquisition, stereo camera calibration was performed using a ChArUco board to extract intrinsic parameters and correct lens distortions later in the workflow. Stereoscopic videos were then recorded in the Pelvitrainer, using the Da Vinci endoscope as during a robotic nephrectomy, moving the camera around the model to maximize surface coverage. Each video lasted approximately one minute.

At this point, enriching frames were extracted from the left of the two video streams at 3 frames per second and manually filtered to exclude blurry or low-quality images.

The AR prototype was launched on the stereoscopic video to extract one initial manual registration keyframe and a variable number of tracking keyframes.

Manual registration was performed in 3D using the Da Vinci console interface, selecting the view where the kidney model was entirely visible and the tumour–parenchyma junction on the kidney surface was best defined. The two video streams were used for the 3D registration, but the keyframes were finally extracted from the left video stream only. After this initial manual registration, the AR prototype tracked the kidney model during camera movements, allowing the operator to select multiple tracking keyframes from different angles. Neural network-based segmentation was disabled to ensure consistency, given the mismatch between this phantom testing environment and the actual surgical environment it was trained for [27, 28].

Figure 3: Experimental setup used for preclinical validation of the AR prototype. From top to bottom and left to right: Da Vinci stereoscopic endoscope, Pelvitrainer box trainer, 3D-printed kidney model coated with AESUB spray, ArUco marker, and calibration checkerboard.

3. Photogrammetric 3D Reconstruction

The AR-derived keyframes, together with the enriching frames previously extracted from the video, were processed using Meshroom (v.2021.1 [29]), an open access tool used to generate photogrammetric models through an SfM pipeline. During this process, the camera calibration parameters that were collected at the beginning of the recording, were synchronised with those used in the AR prototype to ensure consistency.

The resulting SfM models were scaled using CloudCompare [30], an open access tool that gives the ability to visualise and to measure the SfM reconstruction. Knowing that the ArUco marker measures 28 mm, determining its size in the reconstruction allowed us to recover the scale factor converting the 3D model to real-world dimensions.

Subsequently, still within CloudCompare, the area of interest was manually segmented to restrict the zone of extraction for vertex coordinates, which were then used for TRE analysis. This way, the measurement was limited to a clinically relevant region, the tumour–parenchyma junction on the kidney surface.

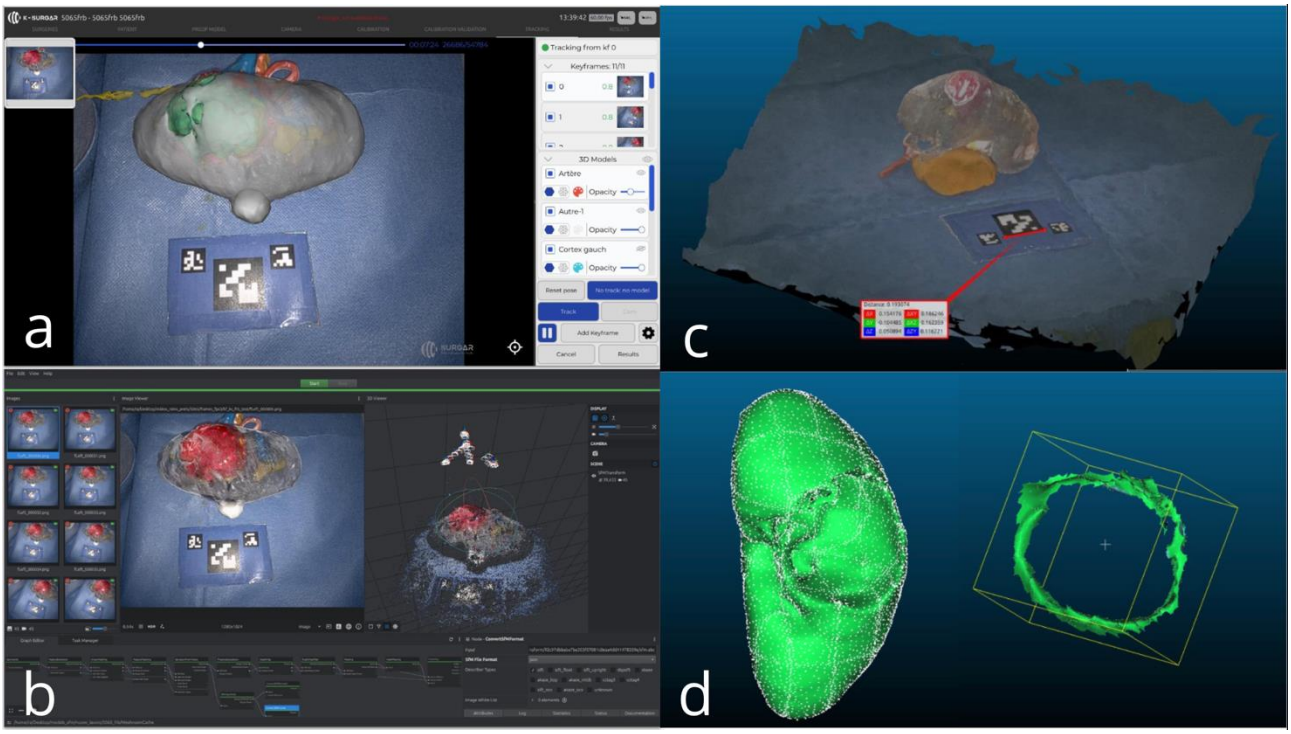

Figure 4: a. manual registration view; b. Photogrammetric reconstruction in Meshroom; c. ArUco marker size measuring on CloudCompare; d. Vertices highlighted in the whole kidney model on the left side of the image and the manually selected tumour-parenchyma intersection on the surface of the kidney on the right side of the image.

4. Target Registration Error Computation

A specific Python script was written and ran for each kidney to calculate the Euclidean distance between corresponding vertices located in the previously selected tumour-parenchyma junction area, from the AR-registered 3D model and the SfM-based ground truth. This calculation was made for every keyframe. TRE, used as the principal accuracy metric, was defined as the average of the TRE values across all vertices in the segmented tumour-parenchyma intersection area and was calculated for each manual and tracking keyframe.

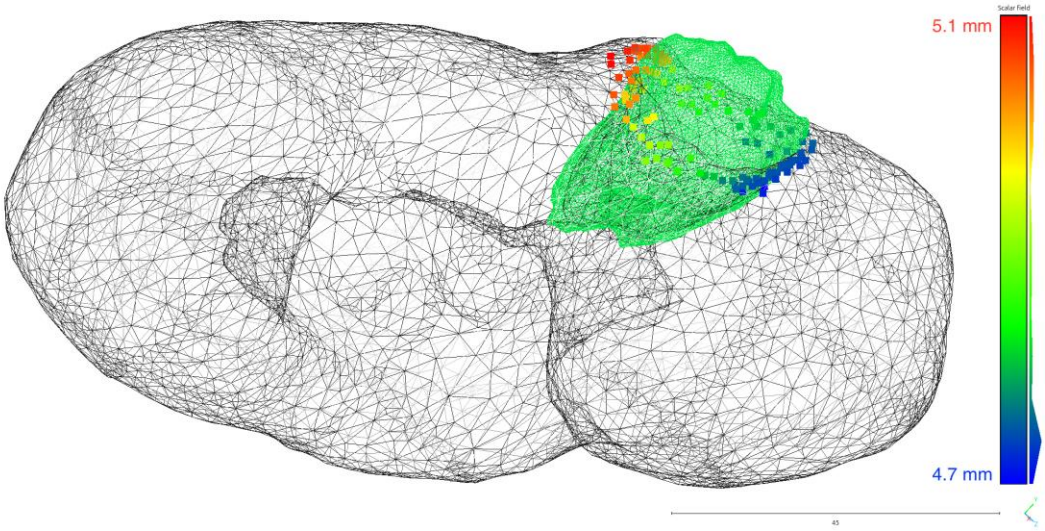

Figure 5: Distribution of the vertices used for TRE calculation, colour-shaded according to their TRE values.

EXPERIMENTAL RESULTS

1. Manual Registration Accuracy

TRE was computed for the manually registered keyframe of each kidney model across three independent operators. Overall, manual TRE values were consistently below 10 mm.
Inter-operator variability was limited, with maximum differences of 5 mm in TRE results between the operators for any given model.
The analysis of inter-operator reproducibility showed that, when considering the average of the three operators, ICC(3,*k*) = 0.80 (95% CI: 0.41–0.95), suggesting a reasonably reproducible manual registration process across users and so indicating good overall reliability of the method according to the classification proposed by Koo & Li (2016).

Kidney_ID	Operator 1 TRE (mm)	Operator 2 TRE (mm)	Operator 3 TRE (mm)
3	5,76	5,04	7,54
7	5,62	6,06	7,35
9	5,50	8,09	5,97
11	4,79	4,49	5,64
13	6,99	5,02	8,54
42	7,28	7,64	5,75
5009	6,32	9,23	4,72
5021	4,22	2,69	3,36
5026	5,60	0,54	2,23
5065	1,98	0,45	2,57
mean	5,41	4,93	5,37

Table 2: TRE values for manual keyframes.

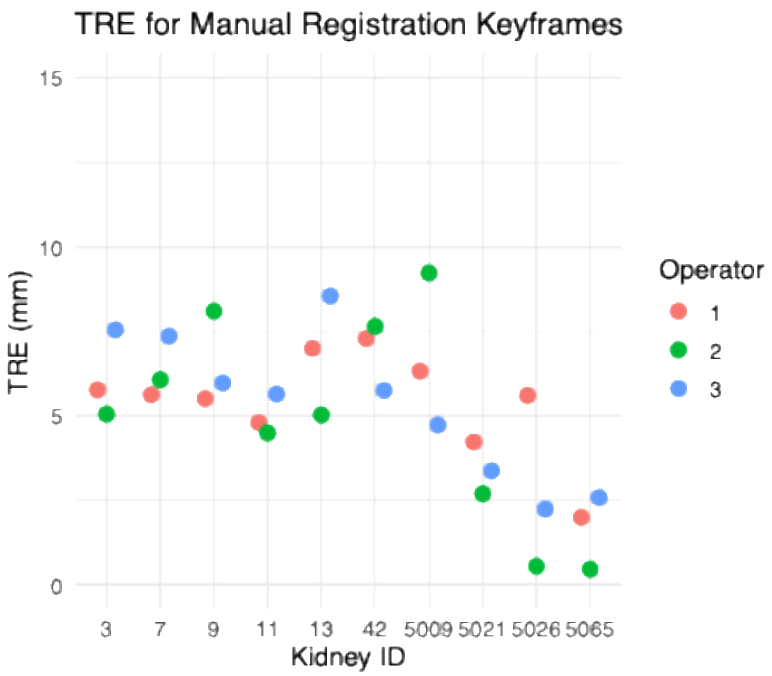

Figure 6: Manual TRE values for each operator and each kidney model.

2. Tracking Accuracy

TRE values for automatically tracked keyframes were consistently below 10 mm also.

Kidney_ID	Operator 1 - TRE (mm)	N KF1	Operator 2 - TRE (mm)	N KF2	Operator 3 - TRE (mm)	N KF3
3	6.73 ± 2.66	11	5.57 ± 2.02	13	3.83 ± 1.47	12
7	7.33 ± 1.36	15	2.92 ± 1.53	14	3.41 ± 1.42	10
9	5.16 ± 1.93	18	6.41 ± 3.03	10	6.48 ± 1.54	10
11	5.08 ± 2.48	13	4.83 ± 2.38	8	7.04 ± 0.8	11
13	6.04 ± 1.09	15	4.68 ± 1.18	11	7.99 ± 0.6	12
42	6.65 ± 1.19	12	7.69 ± 2.1	7	5.61 ± 0.74	10
5009	4.3 ± 1.3	11	7.57 ± 1.73	10	4.91 ± 1.3	11
5021	4.11 ± 0.38	13	2.7 ± 2.24	10	5.12 ± 2.37	11
5026	5.88 ± 1.59	12	0.96 ± 0.26	11	2.3 ± 0.74	12
5065	2.27 ± 0.64	11	1.18 ± 0.56	13	2.18 ± 0.52	10
Mean N KF		13.1		10.7		10,9

Table 3 Mean TRE values and standard deviation for tracking keyframes, with mean number of tracking keyframes per operator.

No significant correlation was found between the number of keyframes and the mean Tracking TRE (r = –0.049, p = 0.8), indicating that tracking accuracy does not depend on the number of keyframes used.

Using a linear mixed-effects model with the printed kidney phantoms as a random effect, the variability attributable to intrinsic differences between kidneys was modest (SD ≈ 1.22 mm), compared with residual variability (SD ≈ 1.59 mm), suggesting that most of the variation in mean tracking TRE is not explained by differences between kidney phantoms.

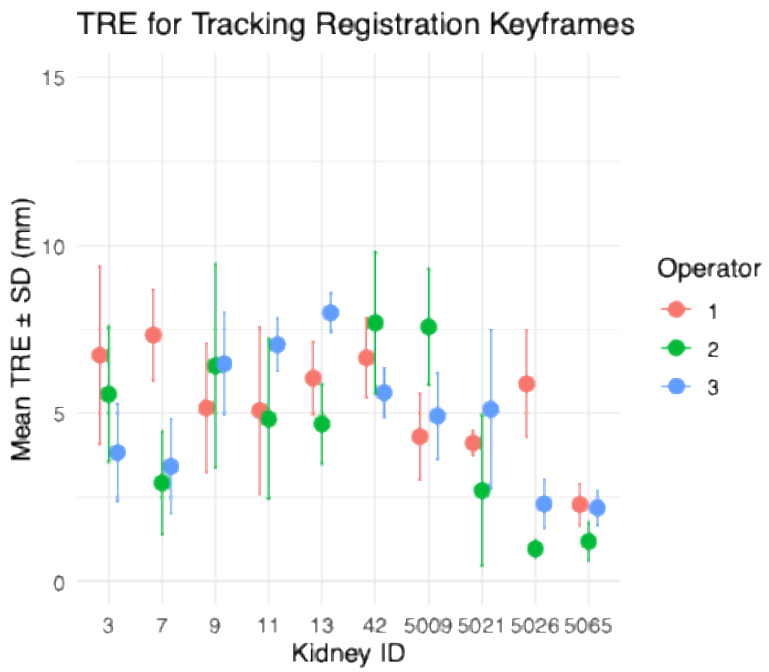

Figure 7: Mean TRE values and standard deviation for each operator and each kidney model.

3. Comparison of Manual TRE and Tracking TRE

Kidney_ID	Mean Δ TRE (Manual – Tracking) (mm)
3	0,74
7	1,79
9	0,51
11	-0,68
13	0,61
42	0,24
5009	1,16
5021	-0,55
5026	-0,26
5065	-0,21

Table 4: Mean precision gain of Tracking TRE compared to corresponding Manual TRE per kidney phantom, in millimetres.

Overall, the mean TRE value was 4.91 mm (± 2.43). The Manual TRE and tracking TRE were comparable, with mean values respectively of 5.23 mm (± 2.24) and 4.88 mm (± 2.45).

A strong positive and statistically significant correlation was found between manual TRE and tracking TRE (β = 0.89; $p < 0.001$), indicating that cases with higher manual TRE tended to also show higher tracking TRE.

We investigated whether tracking was more or less precise compared with the corresponding manual registration TRE. The tracking gain, defined as the improvement in precision achieved by tracking relative to the manual TRE, varied across operators. Overall, the correlation between manual TRE and tracking gain was weak and not statistically significant (β = 0.16; p = 0.093). The mean tracking gain was –0.99% for Operator 1, –16.1% for Operator 2, and 3.46% for Operator 3, with an overall mean gain of –4.54%, indicating that tracking did not consistently improve precision across all operators and could be less precise compared to manual registration.

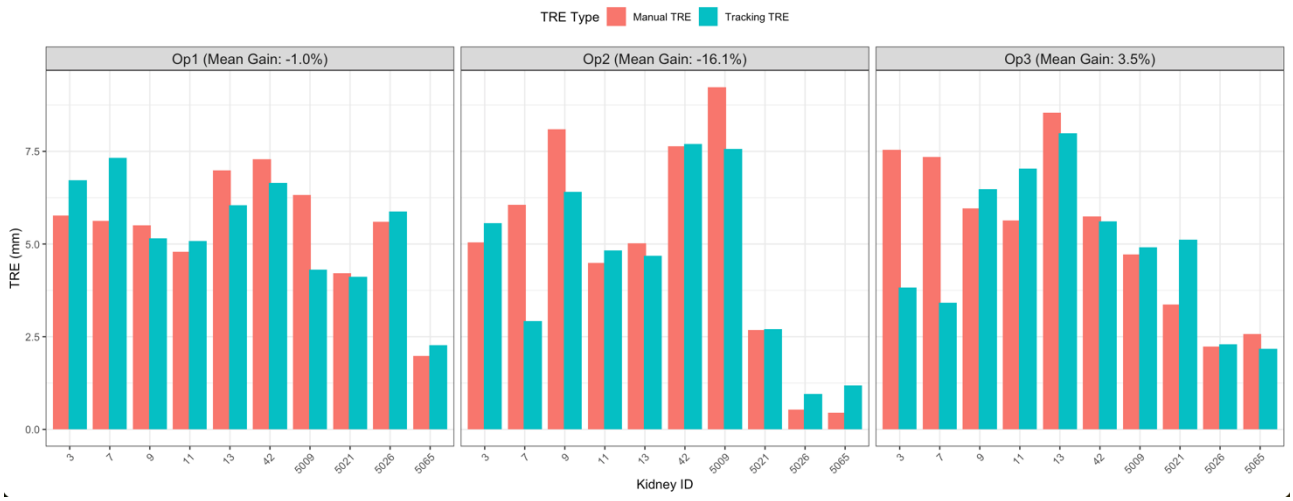

Figure 8: Manual TRE vs Tracking TRE per operator.

4. Registration Time and Correlation with TRE

The average manual registration time across all operators was 173.6 ± 76.4 seconds. Operator 3 was the fastest (144.3 ± 58.1 s), followed by Operator 1 (166.6 ± 51.7 s) and Operator 2 (209.8 ± 101.6 s).

Kidney_ID	Operator 1 Registration time (s)	Operator 2 Registration time (s)	Operator 3 Registration time (s)
3	188	181	130
7	94	48	62
9	109	245	171
11	142	100	99
13	248	166	184
42	121	250	199
5009	144	270	132
5021	213	252	224
5026	216	415	186
5065	191	171	56
Mean	166.6 ± 51.7	209.8 ± 101.6	144.3 ± 58.1

Table 5: Time spent for manual Registration.

No statistically significant correlation was found between registration time and manual TRE for any operator (Operator 1: $r = -0.15$, $p = 0.68$; Operator 2: $r = -0.18$, $p = 0.63$; Operator 3: $r = -0.07$, $p = 0.85$). Weak negative trends were observed, suggesting that a longer registration time might be associated with a slightly better TRE, although this relationship was not statistically significant.

Similarly, no significant correlation was observed between registration time and mean tracking TRE (Operator 1: $r = -0.26$, $p = 0.47$; Operator 2: $r = -0.05$, $p = 0.89$; Operator 3: $r = 0.38$, $p = 0.27$). Just weak trends were observed, with negative correlations for Operators 1 and 2 and a positive correlation for Operator 3, although none reached statistical significance.

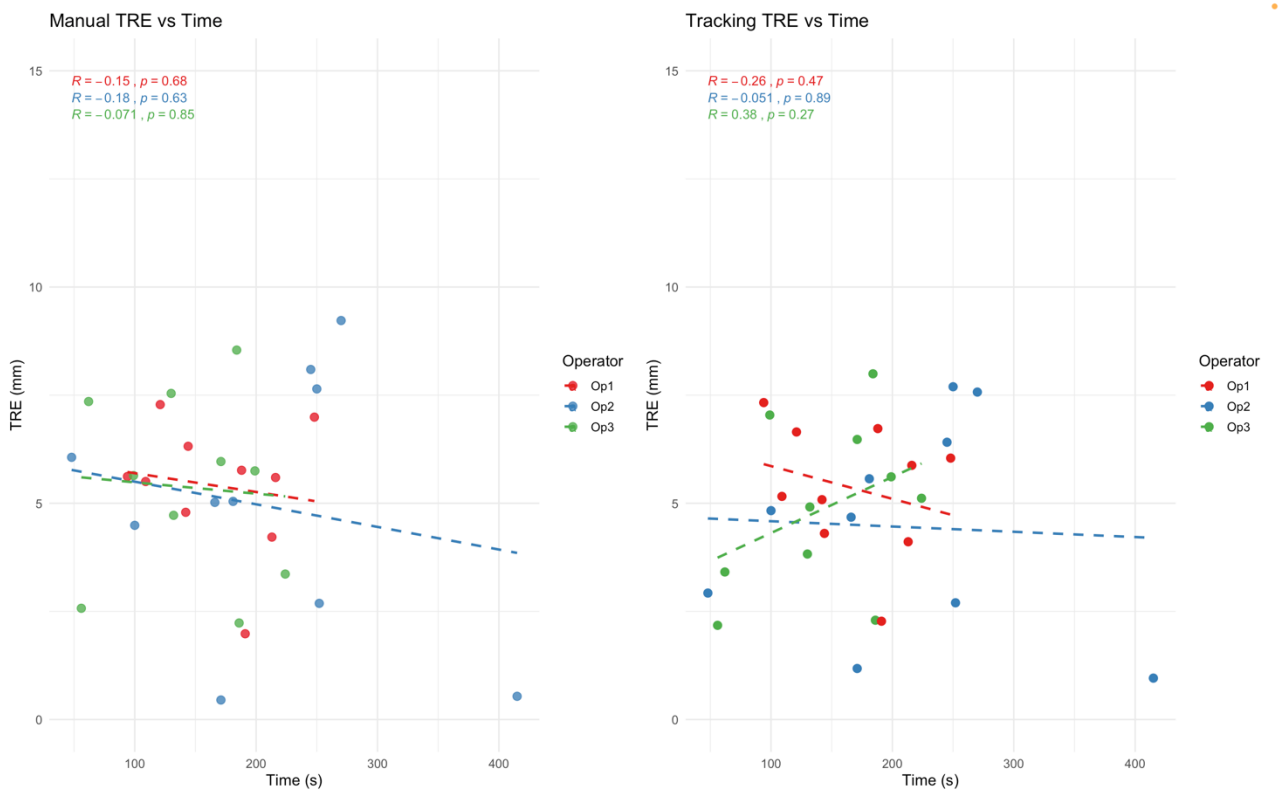

Figure 9: Manual TRE vs Time on the left of the image and Tracking TRE vs Time on the right of the image.

DISCUSSION

In this preclinical study on an AR prototype, the overall mean TRE values were below 5 mm. Both manual and tracking-based registrations consistently achieved TRE values below 10 mm around the tumour–parenchyma junction on the kidney surface, which is a critical anatomical region for nephron-sparing procedures. These objective and quantifiable findings support the reliability and accuracy of the AR prototype. Manual registration showed limited inter-operator variability for the same kidney model. This indicates a robust and repeatable workflow and suggests that identifying anatomical landmarks and registering models in 3D using the TilePro interface, likewise in routine surgical condition, can be performed by different surgeons with a low degree of user-dependent variation.

It should be noted that direct numerical comparisons with previous studies reporting TRE values below 2 mm are not entirely applicable, as those works were conducted using methodologies based on fiducial landmarks, optical trackers, or target organs characterized by a greater number of recognizable features than the kidney [31, 32]. In contrast, our protocol evaluates the performance of an AR prototype applied to 3D-printed renal models designed to realistically reproduce the morphology of an organ with few identifiable surface features, using a workflow integrated into the Da Vinci robotic system without additional trackers. This approach inherently increases measurement variability but enhances clinical transferability. Setting the threshold at 5 mm provides a practical upper limit that accommodates tracking limitations and workflow constraints, while remaining within the range shown by empirical evidence from validation in the medical literature to support accurate surgical navigation and resection.

Contrary to our initial assumption, no significant correlation was found between the mean Tracking TRE and either the number of keyframes (r = –0.049, p = 0.8) or kidney-specific differences, as the variability attributable to phantom anatomy was modest (SD ≈ 1.22 mm). These results suggest that

tracking performance across operators and phantoms is mainly influenced by general reconstruction noise or image acquisition conditions, rather than by the number of keyframes or anatomical variability, and this is consistent with previous work reported in the literature [33, 34]. Therefore, a careful operator-based selection of keyframes remains key to ensuring consistent results.

The analysis showed that tracking and manual registration achieved comparable levels of accuracy, with mean TRE values of 5.23 mm (± 2.24) and 4.88 mm (± 2.45), respectively. The strong positive correlation between manual and tracking TRE suggests that the operator's precision during the initial manual registration phase influences the reliability of the AR content in subsequent steps. Furthermore, the evaluation of tracking gain indicated that tracking did not consistently provide an improvement in precision over manual registration. While one operator achieved a modest positive gain, the other two operators exhibited reduced precision with tracking, resulting in a slightly negative mean gain overall. These findings suggest that tracking performance is operator-dependent and may be influenced by differences in the strategy of keyframe selection, or visual assessment during registration. This highlights the importance of operator experience and acquisition protocols, particularly in cases of complex manual registration, where repeating the process on a more favourable frame may be advisable.

Manual registration time averaged under 3 minutes, marking the feasibility of this part of the workflow in a pre-clinical setup, and eventually for a future intraoperative use. Although no direct correlation was found between registration duration and accuracy, a general trend across the three users suggested that more time spent in the manual registration process might slightly improve its precision. Overall, time alone was not a reliable predictor of performance, while user strategy and system familiarity likely play a larger role.

To be able to evaluate the spatial accuracy of this novel AR prototype in a preclinical context, we developed a specific validation workflow tailored to a controlled, low-variability environment. This empirical measurement approach was specifically designed to assess the system's potential prior to in vivo application and incorporated several novel elements. Above all, the ground truth for TRE calculation was an SfM reconstruction derived from a video of 3D-printed kidney phantoms replicating the virtual models used in the AR prototype. This approach allowed the evaluation of target-point distances between the virtual model displayed by the AR prototype in intraoperative-like video keyframes and the corresponding SfM reconstructions, while avoiding complications such as organ deformation or the presence of blood, smoke, or coagulation in the images, which could affect manual and tracking registration accuracy. While not an absolute reference, the SfM model provided a practical and sufficiently accurate baseline for assessing registration performance.

In addition, a key component of this workflow was the use of a 3D manual registration strategy. By exploiting the stereo video collected with the da Vinci's camera in the da Vinci console's 3D view via the TilePro input, operators performed volumetric registration of the virtual and physical models, rather than relying solely on 2D reprojections. This approach improved depth perception and minimized perspective distortion, ultimately improving registration accuracy.

Several methodological limitations should be considered. First, the SfM model used as reference represents a pseudo–ground truth. Its accuracy depends on the number of images, their resolution, the camera movement stability, the proportion of the object captured in each frame, how much of the frame the object occupies, and the extent of angular coverage. Additionally, the scaling factor was calculated from a manual measurement of the ArUco marker, adding other potential error layers. These steps introduce cumulative uncertainty in the TRE computation. The camera poses associated with each keyframe are inferred retrospectively during the SfM reconstruction process. These estimations may carry implicit errors that are difficult to quantify precisely. The tumour-parenchyma

junction on the kidney surface used for analysis may not perfectly correspond between the virtual and physical models due to segmentation artifacts or 3D printing tolerances. Additionally, in order to compare camera poses from the AR-Prototype system with those from the SfM model, a registration step is required. This step relies on the best possible alignment between the two sets of poses, introducing another potential source of error that is also difficult to isolate and quantify.

That said, although various factors may have negatively influenced the results of this study, the fact that the average TRE value remain below 5 mm supports the fundamental reliability of K-SurgAR-prototype. It is likely that improvements limited to the TRE computation methodology alone could further enhance these already satisfactory results, without requiring changes to the K-SurgAR-prototype system itself.

To improve accuracy and reduce uncertainty in future studies, several methodological improvements would be valuable. The use of external fiducial markers with known coordinates could minimize errors related to manual registration and manual scaling and help for SfM reconstruction accuracy evaluation. Additionally, incorporating high-resolution images from external cameras alongside endoscopic video could improve SfM model quality. Furthermore, future work may integrate stereo video streams, providing a metric scale factor that would help to reconstruct a more reliable ground truth, once calibration and synchronization pipelines are robust enough to avoid introducing new sources of error. Lastly, future studies should explore alternative workflows enabling TRE assessment in other anatomical regions, in order to extend spatial validation to fully endophytic tumours.

An in vivo study, part of an upcoming clinical study, will further assess the AR prototype under realistic surgical conditions. From a practical standpoint, the AR prototype will be integrated into the surgical workflow, with the only difference compared to standard RAPN being the timing of keyframe acquisition and the execution of the initial manual registration. Following this step, the system will be capable of automatically tracking the organ throughout the procedure.

No specific adaptation of the system will be required, as it was primarily developed for this purpose and the neural network responsible for recognizing the kidney surface was trained using RAPN surgical videos. The anticipated challenges mainly relate to difficulties in renal surface recognition, particularly in the presence of bleeding or surgical smoke.

In this setting, TRE will be assessed with automatic calculation of the distance between the virtual tumour boundary and the coagulation mark performed by the surgeon on the kidney surface before the tumour enucleation. This evaluation will help determine whether the system can maintain millimetric precision in the more variable and demanding environment of live surgery.

CONCLUSION

This study presents a novel preclinical methodology to quantitatively evaluate the spatial accuracy of an AR prototype for robot-assisted partial nephrectomy. Using TRE measurements at the tumour–parenchyma junction on virtual models and SfM-based reconstructions of 3D-printed kidney phantoms, the system demonstrated clinically acceptable accuracy with mean TRE values below 5 mm. Manual and tracking-based registrations showed comparable results and low inter-operator variability, indicating robust and reproducible performance. Tracking accuracy was unaffected by keyframe number or anatomical variability, highlighting image quality as the primary determinant of SfM consistency. A forthcoming clinical study will assess in vivo accuracy to support clinical translation of the AR system.

BIBLIOGRAPHY

[1] Campbell SC, Uzzo RG, Karam JA, Chang SS, Clark PE, Souter L. Renal Mass and Localized Renal Cancer: Evaluation, Management, and Follow-up: AUA Guideline: Part II. Journal of Urology 2021;206:209–18. https://doi.org/10.1097/JU.0000000000001912.

[2] Rose TL, Kim WY. Renal Cell Carcinoma: A Review. JAMA 2024;332:1001–10. https://doi.org/10.1001/jama.2024.12848.

[3] Mastroianni R, Chiacchio G, Perpepaj L, Tuderti G, Brassetti A, Anceschi U, et al. Comparison of Perioperative, Functional, and Oncologic Outcomes of Open vs. Robot-Assisted Off-Clamp Partial Nephrectomy: A Propensity Scored Match Analysis. Sensors 2024;24:2822. https://doi.org/10.3390/s24092822.

[4] Casale P, Lughezzani G, Buffi N, Larcher A, Porter J, Mottrie A. Evolution of Robot-assisted Partial Nephrectomy: Techniques and Outcomes from the Transatlantic Robotic Nephron-sparing Surgery Study Group. European Urology 2019;76:222–7. https://doi.org/10.1016/j.eururo.2018.11.038.

[5] Ingels A, Bensalah K, Beauval JB, Paparel P, Rouprêt M, Lang H, et al. Comparison of open and robotic-assisted partial nephrectomy approaches using multicentric data (UroCCR-47 study). Sci Rep 2022;12:18981. https://doi.org/10.1038/s41598-022-22912-8.

[6] Bensalah K, Pignot G, Legeais D, Madec F-X, Lebacle C, Doizi S, et al. Les complications de la néphrectomie totale et de la néphrectomie partielle : quelles sont-elles, comment les prévenir et les prendre en charge ? Progrès en Urologie 2022;32:928–39. https://doi.org/10.1016/j.purol.2022.09.011.

[7] Porpiglia F, Mari A, Bertolo R, Antonelli A, Bianchi G, Fidanza F, et al. Partial Nephrectomy in Clinical T1b Renal Tumors: Multicenter Comparative Study of Open, Laparoscopic and Robot-assisted Approach (the RECORd Project). Urology 2016;89:45–53. https://doi.org/10.1016/j.urology.2015.08.049.

[8] Hung AJ, Cai J, Simmons MN, Gill IS. "Trifecta" in Partial Nephrectomy. Journal of Urology 2013;189:36–42. https://doi.org/10.1016/j.juro.2012.09.042.

[9] Michiels C, Khene Z-E, Prudhomme T, Boulenger De Hauteclocque A, Cornelis FH, Percot M, et al. 3D-Image guided robotic-assisted partial nephrectomy: a multi-institutional propensity score-matched analysis (UroCCR study 51). World J Urol 2021;41:303–13. https://doi.org/10.1007/s00345-021-03645-1.

[10] Yoshitomi KK, Komai Y, Yamamoto T, Fukagawa E, Hamada K, Yoneoka Y, et al. Improving Accuracy, Reliability, and Efficiency of the RENAL Nephrometry Score With 3D Reconstructed Virtual Imaging. Urology 2022;164:286–92. https://doi.org/10.1016/j.urology.2022.01.024.

[11] Checcucci E, Piazza P, Micali S, Ghazi A, Mottrie A, Porpiglia F, et al. Three-dimensional Model Reconstruction: The Need for Standardization to Drive Tailored Surgery. European Urology 2022;81:129–31. https://doi.org/10.1016/j.eururo.2021.11.010.

[12] Amparore D, Piramide F, De Cillis S, Verri P, Piana A, Pecoraro A, et al. Robotic partial nephrectomy in 3D virtual reconstructions era: is the paradigm changed? World J Urol 2022;40:659–70. https://doi.org/10.1007/s00345-022-03964-x.

[13] Amparore D, Sica M, Verri P, Piramide F, Checcucci E, De Cillis S, et al. Computer Vision and Machine-Learning Techniques for Automatic 3D Virtual Images Overlapping During Augmented Reality Guided Robotic Partial Nephrectomy. Technol Cancer Res Treat 2024;23:15330338241229368. https://doi.org/10.1177/15330338241229368.

[14] Piramide F, Kowalewski K-F, Cacciamani G, Rivero Belenchon I, Taratkin M, Carbonara U, et al. Three-dimensional Model–assisted Minimally Invasive Partial Nephrectomy: A

Systematic Review with Meta-analysis of Comparative Studies. European Urology Oncology 2022;5:640–50. https://doi.org/10.1016/j.euo.2022.09.003.
[15] Kim YC, Park C-U, Lee SJ, Jeong WS, Na SW, Choi JW. Application of augmented reality using automatic markerless registration for facial plastic and reconstructive surgery. Journal of Cranio-Maxillofacial Surgery 2024;52:246–51. https://doi.org/10.1016/j.jcms.2023.12.009.
[16] Chu Y, Yang J, Ma S, Ai D, Li W, Song H, et al. Registration and fusion quantification of augmented reality based nasal endoscopic surgery. Medical Image Analysis 2017;42:241–56. https://doi.org/10.1016/j.media.2017.08.003.
[17] Chegini S, Edwards E, McGurk M, Clarkson M, Schilling C. Systematic review of techniques used to validate the registration of augmented-reality images using a head-mounted device to navigate surgery. British Journal of Oral and Maxillofacial Surgery 2023;61:19–27. https://doi.org/10.1016/j.bjoms.2022.08.007.
[18] Fitzpatrick JM, West JB. The distribution of target registration error in rigid-body point-based registration. IEEE Trans Med Imaging 2001;20:917–27. https://doi.org/10.1109/42.952729.
[19] Jackson P, Simon R, Linte C. Effect of uncertainty on target registration error in image-guided renal interventions: from simulation to in-vitro assessment. In: Linte CA, Siewerdsen JH, editors. Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling, Online Only, United States: SPIE; 2021, p. 15. https://doi.org/10.1117/12.2581854.
[20] Moghari MH, Abolmaesumi P. Distribution of Target Registration Error for Anisotropic and Inhomogeneous Fiducial Localization Error. IEEE Trans Med Imaging 2009;28:799–813. https://doi.org/10.1109/tmi.2009.2020751.
[21] Wiles AD, Likholyot A, Frantz DD, Peters TM. A Statistical Model for Point-Based Target Registration Error With Anisotropic Fiducial Localizer Error. IEEE Trans Med Imaging 2008;27:378–90. https://doi.org/10.1109/tmi.2007.908124.
[22] Chegini S, Edwards E, McGurk M, Clarkson M, Schilling C. Systematic review of techniques used to validate the registration of augmented-reality images using a head-mounted device to navigate surgery. British Journal of Oral and Maxillofacial Surgery 2023;61:19–27. https://doi.org/10.1016/j.bjoms.2022.08.007.
[23] Kavoussi NL, Pitt B, Ferguson JM, Granna J, Remirez A, Nimmagadda N, et al. Accuracy of Touch-Based Registration During Robotic Image-Guided Partial Nephrectomy Before and After Tumor Resection in Validated Phantoms. Journal of Endourology 2021;35:362–8. https://doi.org/10.1089/end.2020.0363.
[24] Bernhard J-C, Isotani S, Matsugasumi T, Duddalwar V, Hung AJ, Suer E, et al. Personalized 3D printed model of kidney and tumor anatomy: a useful tool for patient education. World J Urol 2016;34:337–45. https://doi.org/10.1007/s00345-015-1632-2.
[25] Michiels C, Jambon E, Bernhard JC. Measurement of the Accuracy of 3D-Printed Medical Models to Be Used for Robot-Assisted Partial Nephrectomy. American Journal of Roentgenology 2019;213:626–31. https://doi.org/10.2214/AJR.18.21048.
[26] García-Ruiz P, Romero-Ramirez FJ, Muñoz-Salinas R, Marín-Jiménez MJ, Medina-Carnicer R. Fiducial Objects: Custom Design and Evaluation. Sensors 2023;23:9649. https://doi.org/10.3390/s23249649.
[27] Chandelon K, Pitout A, Souchaud M, Desternes J, Margue G, Peyras J, et al. Landmark-free automatic digital twin registration in robot-assisted partial nephrectomy using a generic end-to-end model. Int J CARS 2025;20:1931–40. https://doi.org/10.1007/s11548-025-03473-3.
[28] Khaddad A, Bernhard J-C, Margue G, Michiels C, Ricard S, Chandelon K, et al. A survey of augmented reality methods to guide minimally invasive partial nephrectomy. World J Urol 2022;41:335–43. https://doi.org/10.1007/s00345-022-04078-0.
[29] AliceVision. Meshroom 2021.
[30] GPL Software. CloudCompare 2023.
[31] Jackson P, Merrell K, Simon R, Linte CA. Assessing and reducing registration and targeting uncertainty in video-based image-guided renal navigation: an in vitro study featuring virtual

targets. In: Linte CA, Siewerdsen JH, editors. Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling, San Diego, United States: SPIE; 2022, p. 105. https://doi.org/10.1117/12.2613457.
[32] Zheng Z, Fan J, Shao L, Zhang X, Song H, Ai D, et al. Simultaneous Reconstruction and Tracking for Motion Correction in Image Guided Surgery. IEEE Trans Biomed Eng 2025:1–12. https://doi.org/10.1109/TBME.2025.3614666.
[33] Ketcha MD, De Silva T, Han R, Uneri A, Goerres J, Jacobson MW, et al. Effects of Image Quality on the Fundamental Limits of Image Registration Accuracy. IEEE Trans Med Imaging 2017;36:1997–2009. https://doi.org/10.1109/TMI.2017.2725644.
[34] Jackson P, Simon R, Linte C. Effect of uncertainty on target registration error in image-guided renal interventions: from simulation to in-vitro assessment. In: Linte CA, Siewerdsen JH, editors. Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling, Online Only, United States: SPIE; 2021, p. 15. https://doi.org/10.1117/12.2581854.