ARTICLE TEMPLATE

Kidney Tracking for Live Augmented Reality in Stereoscopic Mini-invasive Partial Nephrectomy

Kilian Chandelon[a,b], Rasoul Sharifian[a], Salomé Marchand[c], Abderrahmane Khaddad[d], Nicolas Bourdel[a,b,e], Nicolas Mottet[c], Jean-Christophe Bernhard[d] and Adrien Bartoli[a,b,f]

[a]Institut Pascal, UMR6602 CNRS, UCA, Clermont-Ferrand University Hospital, France;
[b]SurgAR - Surgical Augmented Reality, Clermont-Ferrand, France;
[c]Department of Urology, Hôpital Nord, Saint-Etienne University Hospital, France;
[d]Department of Urology, Hôpital Pellegrin, Bordeaux University Hospital, France;
[e]Department of Obstetrics and Gynecology, Clermont-Ferrand University Hospital, France;
[f]Department of Clinical Research and Innovation, Clermont-Ferrand University Hospital, France.

ARTICLE HISTORY
Compiled November 23, 2022

ABSTRACT
Mini-invasive partial nephrectomy (MIPN) has many advantages. However, it is challenging for the surgeon to localise the hidden anatomical structures to be spared or resected during surgery. Augmented reality (AR) is a promising localisation assistance approach. Existing AR-MIPN methods augment the endoscopic view with 3D models from preoperative CT registered by means of manual interactions. However, they do not track the kidney in real-time, which considerably reduces usability, as AR is only temporarily available on isolated images. We propose an approach to achieve continuous live AR-MIPN. It uses classical camera calibration and manual initial registration. Its key novelty is a keypoint-based automatic kidney tracking module, with three main technical contributions. First, it performs stereo tracking-by-detection from stereo keyframes, exploiting left-right consistency to maximise robustness. Second, it only considers keypoints within the parenchyma, as segmented by a specifically trained neural network. Third, it improves keypoint detection and matching by a new process that we call stereo perspective correction (SPC). It uses the stereo depth-map and surface flattening to generate an image warp that cancels the perspective effect, improving the performance of keypoint detection and matching. We carried out experiments on semi-synthetic and real surgical datasets to compare several tracking methods, showing that our method outperforms.

KEYWORDS
Stereo tracking; Augmented reality; Robot-assisted partial nephrectomy.

1. Introduction

Partial nephrectomy (PN) is the surgical resection of part of the kidney, usually to remove tumours. PN is nowadays largely performed mini-invasively (MIPN) by laparoscopy or robot-assistance (RAPN) instead of open surgery. MIPN reduces hospital stay, intraoperative blood loss, and operative time. Modern MIPN facilities provide

CONTACT Kilian Chandelon. Email: kilian.chandelon@gmail.com

enhanced 3D vision via a stereo endoscope. However, they challenge the surgeon in the localisation of the kidney's internal anatomical structures, including the endophytic tumours, which are hidden in the parenchyma, and other important internal elements including the arteries, veins and urinary excretory tracts. Consequently, discerning the boundary between tumorous and healthy parenchyma, hence finding the optimal resection plane, can be extremely difficult. The use of intraoperative ultrasound (IOUS) partially mitigates these challenges but remains highly operator-dependent. This owes to the IOUS image being 2D and difficult to mentally re-position in the real surgical field. In other words, defining a resection path in the IOUS image does not directly translate to gesture guidance. A high level of expertise is thus required to achieve optimal endophytic tumour resection while maintaining sufficient tumour margins and sparring as much healthy tissue as possible.

Computer-aided surgery (CAS) has recently emerged as an approach to alleviate the surgical localisation challenges. Such systems make extensive use of augmented reality (AR) to overlay the internal anatomical structures on the live surgical video. The required information is available as a preoperative 3D model reconstructed from the CT scan. AR thus relies on the ability to register the preoperative 3D model on the surgical view and to track the kidney and camera motion to provide the surgeon with accurate and real-time visual augmentations. AR methods have reached impressive results on organs including the uterus [Collins et al. (2021)] and the liver [Espinel et al. (2021)] but still lag behind for the kidney. This is due to three main reasons. First, the kidney visibility is poor, compared to most other organs, owing to the perirenal fat. In practice, only a very small part of the kidney is visible. Second, the kidney is deformable in a global manner. Even if local deformations are limited, the global kidney shape changes during surgery. Third, the kidney undergoes large motion when mobilised by the surgeon.

Existing AR methods for MIPN are essentially limited to the manual overlay of the preoperative 3D model on selected surgical images by an operator [Su et al. (2009); Nakamura et al. (2010); Pratt et al. (2012); Chen et al. (2013); Wang et al. (2015); Porpiglia et al. (2018, 2019); Schiavina et al. (2020); Amparore et al. (2022b)]. While this already benefits surgery, this is far from the potential impact that an automatic real-time AR system would bring. We thus identify automatic, reliable and real-time kidney tracking as the main current technical limitation and challenge towards allowing AR to express its full potential in MIPN.

We propose a kidney stereo tracking approach for live AR in MIPN. Our system combines initial manual registration of the preoperative 3D model to stereo keyframes and a new online keypoint-based automatic kidney tracking module using segmented flattened images. It is the first system to achieve real-time reliable AR on the kidney. We achieve this with three technical contributions. First, we perform tracking-by-detection from stereo keyframes, exploiting left-right consistency and the stereo depth-map to maximise robustness. Tracking-by-detection consists in detecting the organ in each video frame independently, by detecting and matching keypoints to the stereo keyframes. Contrarily to frame-to-frame tracking, which relies on temporal consistency and requires regular manual re-initialisation, this achieves far greater robustness. Second, we only track keypoints within the parenchyma, as segmented by a specifically trained neural network. This reduces the search space, improving performance and reducing computation time. Third, we improve keypoint detection and matching by a process we call stereo perspective correction (SPC). Indeed, keypoint detectors and descriptors such as SIFT degrade performance with viewpoint changes. Our method exploits the stereo depth-map and surface flattening to generate an image

warp that cancels the perspective effect in the stereo images. We evaluate our kidney stereo tracking method on semi-synthetic images and on postoperative surgical patient images. It successfully tracks the kidney despite strong perspective distortions caused by camera motion for which existing tracking methods fail. For view angles between keyframe and current image between 20° and 45°, using SPC increases the number of correct keypoint matches by 10%. Overall, the proposed stereo method increases by 209.8% the number of tracked frames over the best performing monocular method.

2. Related Work

We first focus on surgical AR and then on a technical point related to SPC.

2.1. AR in Surgery

AR systems in surgery are expanding quickly and now address a wide range of surgical specialties. The existing systems are either marker-based or markerless. Using markers may greatly simplify the difficult registration and tracking problems. In orthopaedics and neurosurgery, the existing commercial systems are all marker-based. They enhance the surgical view with CT, MRI or location data. However, markers cannot always be used in mini-invasive surgery (MIS), in particular they cannot in MIPN. Recent work has thus attempted to achieve markerless registration and tracking, which is tremendously more difficult, especially for moving and deforming organs.

2.2. AR in Mini-invasive Abdominal Procedures

The development of AR systems in abdomino-pelvic MIS is an active research field where preoperative 3D models obtained from MRI or CT are used. A real-time system was proposed for the uterus, assuming that its intraoperative motion is rigid [Collins et al. (2021)]. Static systems were proposed for the liver, which extensively deforms [Espinel et al. (2021); Haouchine et al. (2016); Robu et al. (2018); Modrzejewski et al. (2019)]. Navigation systems exploiting marker-based optical endoscopic tracking were proposed for the pancreas [Ieiri et al. (2011); Okamoto et al. (2015); Tang et al. (2021)]. The uterus system [Collins et al. (2021)] was tried on the kidney with limited performance [TeluobGuillaume et al. (2019)], because its registration module depends on the organ silhouette, which is not visible for the kidney in general, and because its tracking module breaks down in the presence of global kidney deformation caused by mobilisation.

2.3. AR in Mini-invasive Partial Nephrectomy

There exists a significant body of research on the development of AR systems for MIPN [Khaddad et al. (2022)]. This research mostly focuses on the registration step, which deforms the preoperative 3D model to match a surgical view. These systems thus can perform AR on isolated surgical images to visualise tumour margins [Amir-Khalili et al. (2013)]. However, they require substantial surgeon interactions. Specifically, the kidney non-fiducial registration techniques are manual rigid [Su et al. (2009); Nakamura et al. (2010); Pratt et al. (2012); Chen et al. (2013); Wang et al. (2015); Porpiglia et al. (2018); Schiavina et al. (2020); Amparore et al. (2022b)], manual non-rigid [Por-

piglia et al. (2019)], surface-based [Altamar et al. (2011); Amparore et al. (2022a)] or point-cloud based [Zhang et al. (2019)]. The research in kidney tracking, required to facilitate real-time continuous AR in spite of organ or camera motion, lags behind. Tracking was attempted in [Su et al. (2009)] with a stereo surface-based technique combined with iterative closest point (ICP). Owing to its dependence on the previous images, this method requires regular manual re-initialisation, strongly harming usability. Tracking was attempted in [Amparore et al. (2022a)] using monocular fluorescence imaging, a technique which is not widely available and requires indocyanine green injection.

2.4. Correction of Perspective Distortions

Several works proposed to remove the perspective distortions prior to feature extraction [Köser and Koch (2007); Cao and McDonald (2009); Zeisl et al. (2013)]. The recent method [Toft et al. (2020)] uses a depth-map generated by the mono-depth network to back-project the pixels to 3D and cluster the surface normals into three dominant directions, identifying planar regions. A homography is eventually computed for each cluster to rectify each planar region and keypoints are detected from the rectified image. A similar solution was proposed in [Zeisl et al. (2012)] to correct the perspective distortion in objects with developable surfaces. After detection and parameterisation of developable surfaces in the depth-map, the selected surfaces are unrolled into planes and texture-mapped. The keypoints are eventually detected in the unrolled images. The current state of art on perspective distortion removal is not readily applicable to complex surfaces, in particular to anatomical structures such as the kidney, that contain curved, non-developable and non-smooth shapes.

2.5. Contributions

We propose the first stereo tracking method applicable to the kidney in MIPN. Technically, our contribution is three-fold. First, our method goes further than [Su et al. (2009)] by using tracking-by-detection as in [Collins et al. (2021)], whilst improving keypoint detection using the epipolar geometry and dual matching. Second, in contrast to [TeluobGuillaume et al. (2019)] and [Amparore et al. (2022a)], our method uses semantic segmentation images to localise the kidney parenchyma, improving tracking performance by specifically focusing on the kidney's motion. Third, contrarily to [Toft et al. (2020); Zeisl et al. (2012)], our SPC method works on general surfaces, including complex anatomical surfaces, improving tracking efficiency.

3. Methodology

3.1. System Overview

Our stereo system pipeline, shown in figure 1, is inspired from the monocular system [Collins et al. (2021)]. The preoperative 3D model is reconstructed from the segmented preoperative CT scan with contrast. First, the early artery and urinary phases are non-rigidly registered. Second, the six anatomical structures (kidney, kidney cortex, urinary tract, renal artery, renal vein and tumour) are extracted using a commercial software (Synapse 3D, Fujifilm) and exported in obj format.

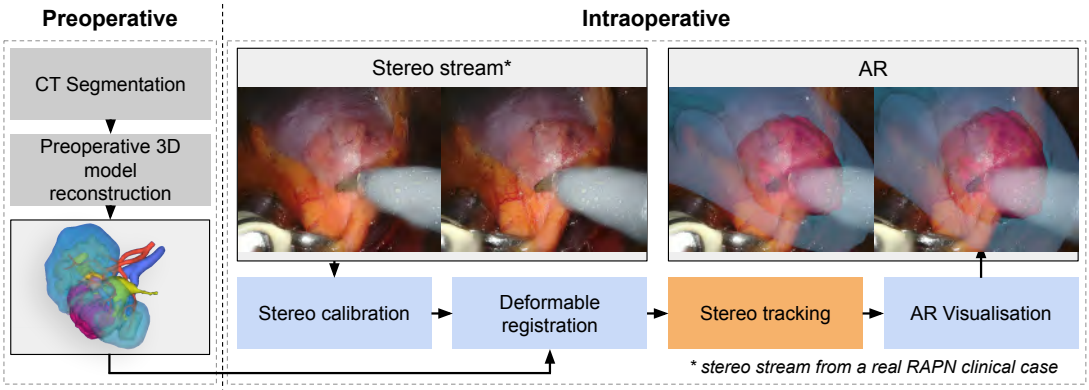

Figure 1. Proposed AR system pipeline, extending [Collins et al. (2021)] to stereo endoscopes in MIPN. The core proposed technical novelties lie in the stereo tracking module.

When surgery starts, the stereo endoscope is intrinsically and extrinsically calibrated automatically using OpenCV and a ChArUco external reference pattern. We then select N stereo images where the kidney is well visible to serve as stereo keyframes, which we manually register to the preoperative 3D model using a simple interactive system, as in [Su et al. (2009); Nakamura et al. (2010); Pratt et al. (2012); Chen et al. (2013); Wang et al. (2015)]. These stereo keyframes are crucial in the tracking process, which works by matching keypoints from the current stereo image to the stereo keyframes in real-time, selecting an optimally matching stereo keyframe. These keypoint matches allow one to estimate the endoscope motion relative to the kidney between the selected stereo keyframe and the current stereo image. Recall that the location of the preoperative 3D model with respect to the stereo keyframes has been established by registration once and for all. Composing this endoscope motion with the selected stereo keyframe registration to the preoperative 3D model finally allows one to transfer any anatomical structure reconstructed in the preoperative 3D model to the current stereo image and realise AR visualisation. We typically use $N \in [5, 10]$ stereo keyframes.

3.2. Tracking Overview

Our stereo tracking system exploits the set of stereo keyframes, called the *stereo keyframe database*, to implement tracking-by-detection. Given the current stereo image from the endoscope, our tracking system works by estimating the organ pose between the coordinate frames of the current stereo image and of one of the stereo keyframes, in real-time. Composing the known transformation between the preoperative 3D model to the stereo keyframe and the computed pose from the stereo keyframe to the current stereo image allows the system to transfer any pre-segmented anatomical element and overlay it during surgery. In this approach, the stereo keyframes form a reference, from which an absolute pose is computed. This avoids the drift which typically occurs through error accumulation in incremental methods based on relative camera pose. The robustness of our system is also stronger than these methods, as it processes each stereo frame independently of the past frames, without requiring a prior initialisation, and does not require any manual re-initialisation in case of temporary failure to track a frame. Specifically, we achieve this principle in two main phases, shown in figure 2, which rely on six components given in section 3.3. A crucial proposed component

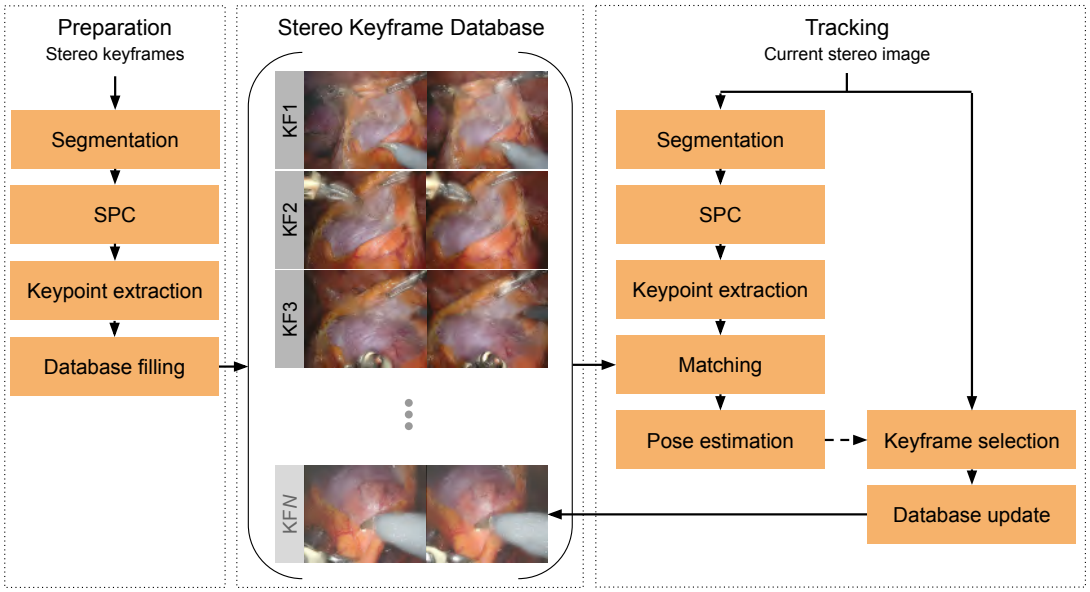

Figure 2. Proposed stereo tracking-by-detection system. In the middle column, KF means keyframe.

is SPC, which warps a given stereo image to a new stereo image. SPC standardises the images in a way that largely cancels the effects of camera projection, including perspective distortion.

3.2.1. Phase 1: Preparation

The N stereo keyframes are stereo images registered to the preoperative 3D model. This means that 3D transformations are available that map any point from the preoperative 3D model to the stereo keyframe coordinates. We prepare each stereo keyframe to subsequent tracking-by-detection by following three steps for left-right images: 1) we automatically segment the kidney parenchyma, 2) we apply SPC and 3) we detect SIFT keypoints in the parenchyma region and store them in the *stereo keyframe database*.

3.2.2. Phase 2: Tracking

In tracking-by-detection, each stereo image, which is composed of its left and right images, is processed independently. We start by applying the same three steps as in the stereo keyframe preparation phase. We follow them by three extra steps: 4) we match the left and right image SIFT keypoints from the current stereo image to each of the stereo keyframes, under stereo epipolar consistency, and select the best matching stereo keyframe, 5) we compute camera absolute pose from the keypoint matches and compose it with the stereo keyframe registration to the preoperative 3D model and 6) we update the *stereo keyframe database* with the current stereo frame, if required.

3.3. Preparation and Tracking Components

3.3.1. Deep Semantic Segmentation of the Kidney Parenchyma

This component aims at providing the probability of each pixel to belong to the kidney parenchyma in the *stereo keyframe database* and the current stereo frames. Segmentation defines the search area for precise tracking but is not sufficient on its own to achieve AR. Indeed, segmenting only cannot be used to relate the surgical images to the preoperative information at the pixel level; hence, it does not convey the precise location where the virtual information must be augmented and does not make AR from preoperative images possible. To perform automatic kidney parenchyma segmentation, we collected and annotated 5091 images from 5 RAPN surgeries in two hospitals to create our dataset. The annotation rules defined in collaboration with clinical experts take into account the accessibility problem of the renal parenchyma during MIPN. The renal surface is enclosed by the renal fascia and under perirenal fat. Importantly, the segmentation should only contain the clean surface of the renal parenchyma, without including fascia, bleeding or any separable tissue, to guarantee reliable kidney tracking. In particular, the fascia is a transparent membrane whose motion may be different of the kidney's. We applied data augmentation with standard image transformations, namely rotation with a random angle chosen between -25° and 25°, white Gaussian noise with standard deviation 0.4 (about 0.16% of the maximal image intensity), uniform random Gaussian blur with a kernel standard deviation within $[1; 1.5]$ px and horizontal flipping. The parameters are chosen to match realistic occurrences. The total number of images available for training is 25456. We trained a UNet [Ronneberger et al. (2015)] using a cross-validation strategy with 5 folds to reduce evaluation bias and a DICE loss. To take into account the temporal ordering of the stereo stream, we additionally use a pre-trained space-time memory network (STM) [Oh et al. (2019)] to propagate the segmentation mask over 100 frame windows. A morphological dilation operation with a 3x3 rectangular kernel is finally applied to the predicted masks.

3.3.2. Stereo Perspective Correction

This component aims at warping the stereo keyframes in the database and the current stereo images in a way that correct the effect of perspective. This is done by the proposed Stereo Perspective Correction (SPC) method, shown in figure 3. The goal of SPC is to improve keypoint matching. We take advantage of stereo vision to recover the 3D shapes. As input, our method uses the stereo image's depth-map to reconstruct the 3D point cloud (figure 3.C) based on the stereo camera parameters. The point cloud is obtained by computing a filtered disparity map (figure 3.B) using semi-global block matching (SGBM) [Hirschmüller (2008)] and fast global image smoothing based on weighted least-squares filtering (WLS) [Min et al. (2014)]. In parallel, the kidney is segmented in the images (figure 3.G). Then, an intraoperative 3D model is obtained by triangulating (triangular meshing) the points in the 3D point cloud within the organ mask (figure 3.D). The triangulated intraoperative 3D model approximates a differentiable surface. We can thus use Riemann's mapping theorem, which guarantees that conformal maps always exist, and map the model to $\mathbb{R}^2$ with minimal angular distortion [Hormann et al. (2007)]. We use least-squares conformal mapping (LSCM) [Lévy et al. (2002); Mullen et al. (2008)] implemented in the libigl C++ library (figure 3.E), which was originally developed to find a 2D parameterisation of a given 3D mesh. It flattens the 3D model whilst approximately preserving the angles. Concretely, it solves an optimisation problem searching for the transformation that

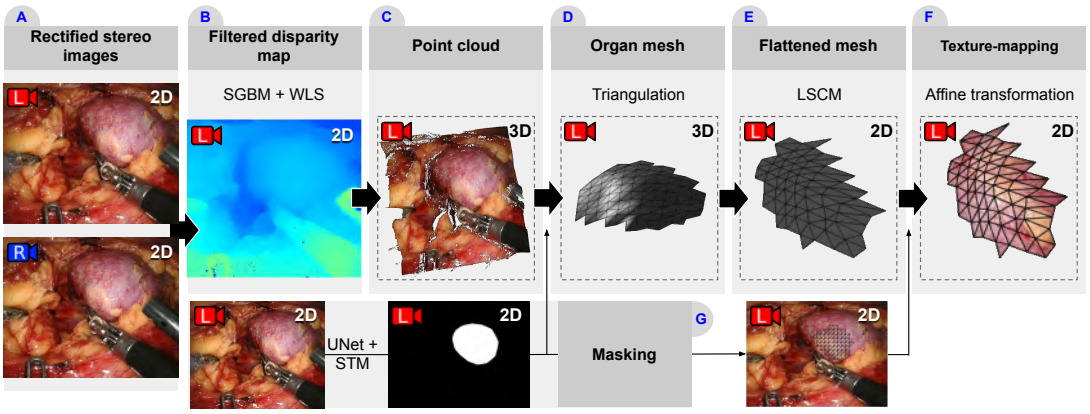

Figure 3. Pipeline of the proposed SPC method applied to the left image; the same pipeline is also run on the right image to obtain two flattened images. A) Rectified current stereo images. B) Filtered disparity map computed with SGBM and WLS. C) Point cloud generated from the disparity map converted into a depth-map. D) Intraoperative 3D model obtained by triangular meshing. E) Flattened 2D model produced by LSCM. F) Flattened 2D model texture-mapped with piecewise affine image warping. G) Kidney segmentation mask from the UNet and STM networks.

maps triangles from $\mathbb{R}^3$ to $\mathbb{R}^2$ while minimising conformal distortion and preserving triangle connectivity. Eventually, the flattened model is texture-mapped by warping the original images (figure 3.F) using a piecewise affine transformation defined by the vertex correspondences. As the surface typically has low curvature relative to the mesh resolution, the texture transformation between the scene plane and the image is locally well approximated by an affine transformation. Therefore, we can use local barycentric coordinates to find the corresponding pixel positions between the flattened triangles and the source image to texture-map the flat model. We warp both the left and right images and thus obtain two flattened images.

3.3.3. Stereo Keypoint Detection

This component aims at detecting keypoints in the *stereo keyframe database* and current stereo images. The perspective corrected images are pre-processed before keypoint detection to enhance detection performance, as follow: only the green channel of the stereo RGB images is kept and a Gaussian blur with a kernel size of 15×15 and a standard deviation of 2.6 px is applied. These pre-processing steps were chosen empirically. Then, we run a GPU implementation of SIFT [Griwodz et al. (2018)] on each on the two images. In the context of laparoscopic surgery, SIFT provides a more stable estimate than other non-deep learning based detectors [Collins et al. (2021)] and allows one to process full high definition images (1920×1080 px) in less than 10 milliseconds. In order to select optimal matchable keypoints, which can lead to good stereo matches and tracks, we perform brute force matching (BFM) from OpenCV between the left and right images and keep only those keypoints that fulfil the epipolar geometry constraint (EGC), with a tolerance in terms of distance to epipolar lines of 0.07% of the image diagonal, equivalent to 1.5 px for 1920×1080 px images.

3.3.4. Matching and Closest Keyframe Selection

This component aims at matching the keypoints between the current stereo image and the stereo keyframes in the database. Correct stereo matches must comply with two

conditions: 1) the keypoints must meet the EGC between the left and right images, which is checked during stereo keypoint detection, and 2) the descriptors must be in strong agreement, as measured by Lowe's ratio test (LRT). The selected keypoints on the left and right of the current stereo image are matched with the *stereo keyframe database*. This is done independently for the left and the right images, using BFM and LRT. The above matching step is run N times, to match the current stereo image with each of the N stereo keyframes from the database, and the closest stereo keyframe is finally selected as the keyframe with the highest number of matched keypoints.

3.3.5. Pose Estimation

This component aims at estimating the rigid body motion that relates the selected database stereo keyframe and the current stereo images. From the set of correct stereo matches, we find the best camera pose that explains these matches using PnP RANSAC from OpenCV. This hypothetical pose is refined by computing a PnP only on the inliers matches. For the cases where the organ visibility is low, for instance when the organ is out of the field of view, the camera pose cannot be estimated. We detect these cases from the number of inliers found by RANSAC: the pose is rejected and the organ declared invisible if the number of inliers is lower than a threshold of 8 points. Finally, the estimated pose is fed into an extended Kalman filter to cancel jitter.

3.3.6. Updating the Stereo Keyframe Database

This component aims at updating the keyframe database to cope with dynamical organ shape and appearance changes. During surgery, these changes may result in tracking failures due to a lack of relevant information in the stereo keyframe database. Adding stereo keyframes to the database copes with changes occuring between the stereo keyframe database and the current stereo image in terms of organ appearance and 3D orientation. We re-implemented the keyframe database system described in [Collins et al. (2021); Chandelon and Bartoli (2022)] and extended it to stereo images.

4. Experimental Results

4.1. Deep Semantic Segmentation of the Kidney Parenchyma

We trained the UNet for 20 epochs with a batch size of 4 and a learning rate of 0.001. We use stochastic gradient descent (SGD) and a stepwise reduced learning rate scheduler. We evaluated the effectiveness of UNet and STM in segmenting the renal parenchyma on respectively a test dataset of 25 images and 5 sequences of 100 frames from a new RAPN surgery not included in the training images, as shown in figure 4. The evaluation data were randomly selected at the surgical phase when AR may be used, specifically during fat removal, tumour exposure and resection.

We obtained for the UNet an average DICE score of 0.8911 and an average IoU score of 0.8101 and for the STM an average DICE score of 0.9647 and an average IoU score of 0.9431. Our results show that intraoperative segmentation of the kidney parenchyma is a very difficult task. First, the inter-patient appearance variability is large. Second, the intra-patient appearance variability during surgery is large too, owing to manipulation and dissection. Third, precisely localising the boundaries of the kidney parenchyma is difficult even for the experts, due to fuzzy edges, occlusions

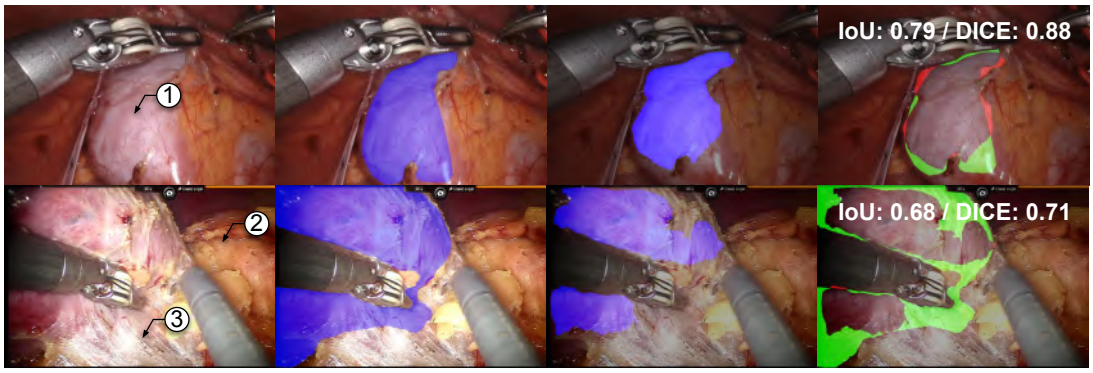

Figure 4. Two test images for semantic kidney segmentation. (left to right) original image, ground truth, UNet prediction, difference between ground truth and prediction (green: addition / red: deletion). Specific anatomical elements: 1- kidney parenchyma, 2- perirenal fat, 3- adhesion.

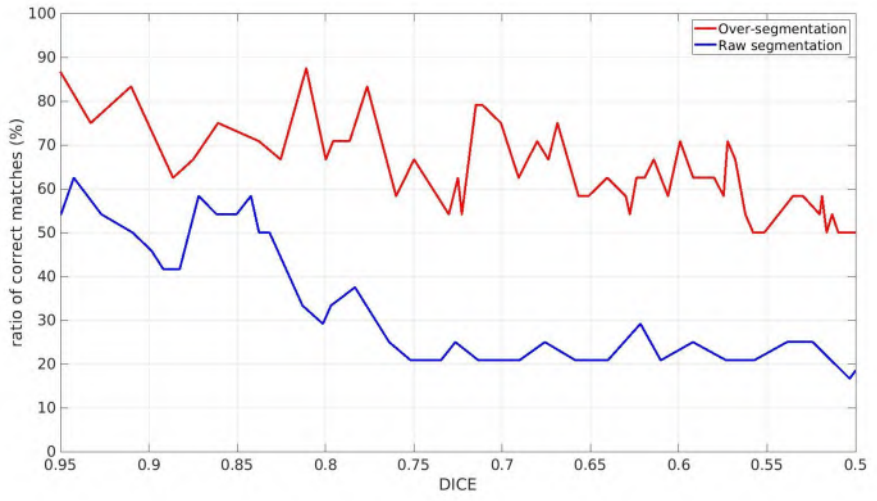

Figure 5. Evaluation of the impact of segmentation accuracy on keypoint detection and matching.

and bleeding.

We evaluated the impact of segmentation accuracy on keypoint detection and matching performance by artificially degrading the segmentation quality. We compared the raw segmentation results and over-segmentation, which is obtained by a mere dilation with a 5x5 rectangular kernel of the raw segmentation. As shown in figure 5, the accuracy of over-segmentation does not have a significant impact on the detection and matching process. This is because segmentation is only used to define a search area.

4.2. Stereo Perspective Correction

We evaluated our SPC method by detecting and describing the SIFT keypoints in the warp images and comparing them with the keypoints extracted from the original images. Matches are computed between a single keyframe considered at position 0° and several endoscope images taken during camera motion. To make the comparison fair, we considered only the keypoints that within the kidney segmentation for both meth-

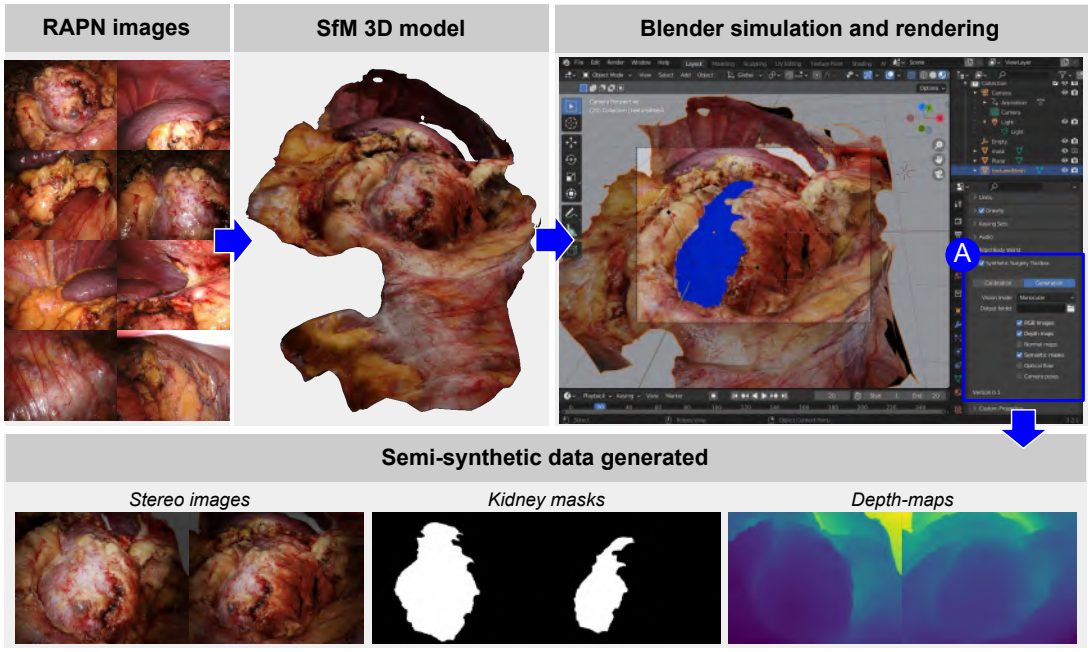

Figure 6. Proposed semi-synthetic data generation from real RAPN images. A- Blender addon specifically designed for stereo images and ground truth MIPN data generation.

ods. We ran two experiments, one on semi-synthetic images and one on real kidney surgery images. The semi-synthetic images are generated from a 3D model reconstructed by structure-from-motion (SfM) using Meshroom from a real RAPN surgery as shown in 6. The stereo endoscope movements are simulated in the computer graphics software Blender, to create different camera viewpoint angles. This 3D scenario allows us to use ground truth data, including stereo camera parameters, endoscope pose, depth-maps and semantic segmentation masks.

The results on semi-synthetic images are shown in the upper part of figure 7. As camera angle increases, perspective distortion becomes noticeable and matching performance significantly decreases. Both the ratio and number of correct matches demonstrate an increasing matching performance with SPC for relative camera angles larger than 20°, figure 7.A and 7.B. The original stereo images matched better for relative camera angles lower than 20°. This is because image warping slightly modifies the image signal, which perturbs SIFT, although the invariance of this detector remains effective for small viewpoint changes. However, as the camera angle increases, the perspective distortion gradually dominates. We have evaluated our framework with SuperGlue [Sarlin et al. (2020)]. The matching performance is on par with SIFT, but the runtime is on average 6 times slower, with a framerate lower than 15 FPS, which is incompatible with real-time processing in surgery.

The use of SPC is fundamental and represents a major gain in complex matching cases with an angle greater than 20°. For low-angle cases, whether SPC is used or not does not impact the success of matching as the number of keypoints is already high. In addition, as shown in figure 7.C, describing features in the flattened domain decreases the descriptor distance between corresponding keypoints. It shows that SPC improves the invariance of feature matching to perspective distortion, as the warped images are more similar than the original images. In other words, on average, SPC improves the discrimination and robustness of SIFT descriptors. The same experiments

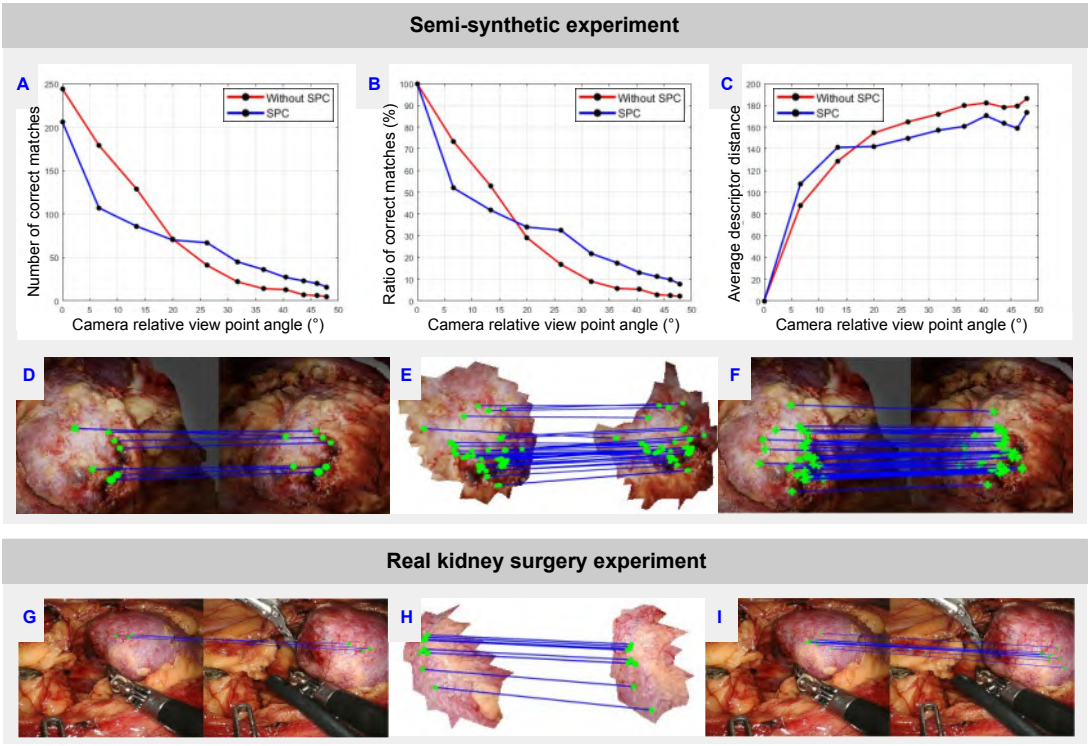

Figure 7. Evaluation of SPC on semi-synthetic and real images. Statistics on semi-synthetic images: A) number of correct matches with respect to relative camera angle, B) ratio of correct matches with respect to relative camera angle, C) average SIFT descriptor distance with respect to relative camera angle. Semi-synthetic example with a relative camera angle of 43.7°: D) correct 7 matches (out of 10 matches) obtained without SPC, E) correct 31 matches (out of 31 matches) in the flat domain with SPC and F) same as E) on the original images. Results on real images: G) correct 3 matches (out of 4 matches) without SPC, H) correct 10 matches (out of 10 matches) in the flat domain with SPC and I) same as H) on the original images.

are performed on real images, the results are shown in the bottom part of figure 7. For large camera angles between the keyframe and the current image, SPC increases the number of correct matches by more than three times. Overall, we observe in both semi-synthetic and real images that SPC slightly decreases matching performance in the easy cases but substantially boosts matching performance in the difficult cases.

4.3. Registration and Tracking

We evaluated our stereo tracking method using an ablation study of the two proposed steps of kidney segmentation and constrained left-right (LR) matching, and a comparison of stereoscopic and monocular tracking. We used three clinical RAPN cases from the Da Vinci (Intuitive Surgical, Sunnyvale, California, United-States) surgical robotic system equipped with a 30° 3D endoscope. We performed stereo calibration based on a pinhole model and rectified the stereo frames. We performed these studies on an Ubuntu PC with Intel Core™ i7-10870H CPU and Nvidia RTX 2060 GPU. The results shown in table 1 are averages for the three clinical cases. We observe that using stereo increases the number of tracked frames by 62.94 percentage points compared to monocular tracking, which is an increase of 209.8%. Using kidney segmentation improves robustness, scaling the standard deviation of the number of matches by a factor of about 7. Without segmentation, keypoint detection is done on the whole image. The

Table 1. Ablation and comparative study statistics.

Baseline	Kidney segmentation	LR matching	Avg. nb. matches	Tracked frames (%)	FPS
Stereo	✗	✗	91 ± 100.62	62.16	84.61
Stereo	✗	✓	47 ± 25.90	92.94	86.29
Stereo	✓	✗	16 ± 13.41	59.35	27.35
Stereo	✓	✓	27 ± 14.91	66.57	28.16
Mono	✗	✗	45 ± 51.10	30.00	150.62
Mono	✓	✗	9 ± 6.93	18.65	36.10

number of keypoints and inliers is higher, the frame is considered to be tracked but this does not mean that the frame is "well" tracked because it is strongly influenced by structures outside the kidney surface, mainly the abdominal cavity. It however does not increase the ratio between the total number of frames and the number of frames tracked.

Using LR matching greatly increases the number of tracked frames, especially when kidney segmentation is not used. Lastly, we observe that, while the monitored run time is compatible with a real usage for all options, kidney segmentation is by far the most expensive component.

5. Conclusion

We have proposed a comprehensive pipeline to realise live AR in MIPN. Our pipeline takes full advantage of stereoscopy to establish robust matches without interference from the surgical field, including the surrounding tissue and instruments. In future work, we will focus on *(i)* extending and diversifying our segmented kidney parenchyma dataset to improve segmentation accuracy, *(ii)* fully integrating our SPC method and stereo to reduce warping artefacts by combining the left and right images, *(iii)* exploiting the keypoints detected with and without SPC in a hybrid method, taking advantage of their complementarity, *(iv)* including non-rigid registration to recover kidney deformations and *(v)* evaluating the impact of our tracking method in terms of usability.

Disclosure statement

The authors report there are no competing interests to declare.

Acknowledgments.

This work was partly funded by Cancéropôle CLARA within the AIALO project.

References

Altamar HO, Ong RE, Glisson CL, Viprakasit DP, Miga MI, Herrell SD, Galloway RL. 2011. Kidney deformation and intraprocedural registration: a study of elements of image-guided kidney surgery. Journal of endourology. 25 3:511–7.

Amir-Khalili A, Nosrati MS, Peyrat JM, Hamarneh G, Abugharbieh R. 2013. Uncertainty-encoded augmented reality for robot-assisted partial nephrectomy: A phantom study. In: AE-CAI.

Amparore D, Checcucci E, Piazzolla P, Piramide F, Cillis SD, Piana A, Verri P, Manfredi M, Fiori C, Vezzetti E, et al. 2022a. Indocyanine green drives computer vision based 3d augmented reality robot assisted partial nephrectomy: the beginning of "automatic" overlapping era. Urology.

Amparore D, Piramide F, Pecoraro A, Verri P, Checcucci E, Cillis SD, Piana A, Busacca G, Manfredi M, Fiori C, et al. 2022b. Identification of recurrent anatomical clusters using three-dimensional virtual models for complex renal tumors with an imperative indication for nephron-sparing surgery: New technological tools for driving decision-making. European Urology Open Science. 38:60 – 66.

Cao Y, McDonald JB. 2009. Viewpoint invariant features from single images using 3d geometry. 2009 Workshop on Applications of Computer Vision (WACV):1–6.

Chandelon K, Bartoli A. 2022. Tracking better, tracking longer: automatic keyframe selection in model-based laparoscopic augmented reality. International Journal of Computer Assisted Radiology and Surgery. 17:1507 – 1511.

Chen Y, Li H, Wu D, Bi K, Liu C. 2013. Surgical planning and manual image fusion based on 3d model facilitate laparoscopic partial nephrectomy for intrarenal tumors. World Journal of Urology. 32:1493–1499.

Collins T, Pizarro D, Gasparini S, Bourdel N, Chauvet P, Canis M, Calvet L, Bartoli A. 2021. Augmented reality guided laparoscopic surgery of the uterus. IEEE Transactions on Medical Imaging. 40:371–380.

Espinel Y, Calvet L, Botros K, Buc E, Tilmant C, Bartoli A. 2021. Using multiple images and contours for deformable 3d-2d registration of a preoperative ct in laparoscopic liver surgery. In: MICCAI.

Griwodz C, Calvet L, Halvorsen P. 2018. Popsift: a faithful sift implementation for real-time applications. Proceedings of the 9th ACM Multimedia Systems Conference:415–420.

Haouchine N, Roy F, Untereiner L, Cotin S. 2016. Using contours as boundary conditions for elastic registration during minimally invasive hepatic surgery. 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS):495–500.

Hirschmüller H. 2008. Stereo processing by semiglobal matching and mutual information. IEEE transactions on pattern analysis and machine intelligence. 30 2:328–41.

Hormann K, Polthier K, Sheffer A. 2007. Mesh parameterization: theory and practice.

Ieiri S, Uemura M, Konishi K, Souzaki R, Nagao Y, Tsutsumi N, Akahoshi T, Ohuchida K, Ohdaira T, Tomikawa M, et al. 2011. Augmented reality navigation system for laparoscopic splenectomy in children based on preoperative ct image using optical tracking device. Pediatric Surgery International. 28:341–346.

Khaddad A, Bernhard JC, Margue G, Michiels C, Ricard S, Chandelon K, Bladou F, Bourdel N, Bartoli A. 2022. A survey of augmented reality methods to guide minimally invasive partial nephrectomy. World Journal of Urology:1 – 9.

Köser K, Koch R. 2007. Perspectively invariant normal features. 2007 IEEE 11th International Conference on Computer Vision:1–8.

Lévy B, Petitjean S, Ray N, Maillot J. 2002. Least squares conformal maps for automatic texture atlas generation. Proceedings of the 29th annual conference on Computer graphics and interactive techniques.

Min D, Choi S, Lu J, Ham B, Sohn K, Do MN. 2014. Fast global image smoothing based on weighted least squares. IEEE Transactions on Image Processing. 23:5638–5653.

Modrzejewski R, Collins T, Seeliger B, Bartoli A, Hostettler A, Marescaux J. 2019. An in vivo porcine dataset and evaluation methodology to measure soft-body laparoscopic liver registration accuracy with an extended algorithm that handles collisions. International Journal of Computer Assisted Radiology and Surgery. 14:1237–1245.

Mullen P, Tong Y, Alliez P, Desbrun M. 2008. Spectral conformal parameterization. Computer Graphics Forum. 27.

Nakamura K, Naya Y, Zenbutsu S, Araki K, Cho S, Ohta S, Nihei N, Suzuki H, Ichikawa T, Igarashi T. 2010. Surgical navigation using three-dimensional computed tomography images fused intraoperatively with live video. Journal of endourology. 24 4:521–4.

Oh SW, Lee JY, Xu N, Kim SJ. 2019. Video object segmentation using space-time memory networks. 2019 IEEE/CVF International Conference on Computer Vision (ICCV):9225–9234.

Okamoto T, Onda S, Yasuda J, Yanaga K, Suzuki N, Hattori A. 2015. Navigation surgery using an augmented reality for pancreatectomy. Digestive Surgery. 32:117 – 123.

Porpiglia F, Checcucci E, Amparore D, Piramide F, Volpi G, Granato S, Verri P, Manfredi M, Bellin A, Piazzolla P, et al. 2019. Three-dimensional augmented reality robot-assisted partial nephrectomy in case of complex tumours: A new intraoperative tool overcoming the ultrasound guidance. European urology.

Porpiglia F, Fiori C, Checcucci E, Amparore D, Bertolo R. 2018. Hyperaccuracy three-dimensional reconstruction is able to maximize the efficacy of selective clamping during robot-assisted partial nephrectomy for complex renal masses. European urology. 74 5:651–660.

Pratt P, Mayer EK, Vale J, Cohen D, Edwards E, Darzi A, Yang GZ. 2012. An effective visualisation and registration system for image-guided robotic partial nephrectomy. Journal of Robotic Surgery. 6:23–31.

Robu MR, Ramalhinho J, Thompson SA, Gurusamy KS, Davidson BR, Hawkes DJ, Stoyanov D, Clarkson MJ. 2018. Global rigid registration of ct to video in laparoscopic liver surgery. International Journal of Computer Assisted Radiology and Surgery. 13:947 – 956.

Ronneberger O, Fischer P, Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation. ArXiv. abs/1505.04597.

Sarlin PE, 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):4937–4946.

Schiavina R, Bianchi L, Chessa F, Barbaresi U, Cercenelli L, Lodi S, Gaudiano C, Bortolani B, Angiolini A, Bianchi FM, et al. 2020. Augmented reality to guide selective clamping and tumor dissection during robot-assisted partial nephrectomy: A preliminary experience. Clinical genitourinary cancer.

Su LM, Vagvolgyi BP, Agarwal R, Reiley CE, Taylor RH, Hager GD. 2009. Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3d-ct to stereoscopic video registration. Urology. 73 4:896–900.

Tang R, Yang W, Hou Y, Yu L, Wu G, Tong X, Yan J, Lu Q. 2021. Augmented reality-assisted pancreaticoduodenectomy with superior mesenteric vein resection and reconstruction. Gastroenterology Research and Practice. 2021.

TeluobGuillaume, CalvetLilian, EspinelYamid, SavareuxLaurent, GuandalinoMarlene, RavelAnne, BourdelNicolas, ChauvetPauline, GuyLaurent, ChabrotPascal, et al. 2019. Preliminary trial of augmented reality performed on a regular and a robot-assisted laparoscopic partial nephrectomies. Videourology.

Toft C, Turmukhambetov D, Sattler T, Kahl F, Brostow GJ. 2020. Single-image depth prediction makes feature matching easier. In: ECCV.

Wang D, Zhang B, Yuan X, Zhang X, Liu C. 2015. Preoperative planning and real-time assisted navigation by three-dimensional individual digital model in partial nephrectomy with three-dimensional laparoscopic system. International Journal of Computer Assisted Radiology and Surgery. 10:1461–1468.

Zeisl B, Köser K, Pollefeys M. 2012. Viewpoint invariant matching via developable surfaces. In: ECCV Workshops.

Zeisl B, Köser K, Pollefeys M. 2013. Automatic registration of rgb-d scans via salient directions. 2013 IEEE International Conference on Computer Vision:2808–2815.

Zhang X, Wang J, Wang T, Ji X, Shen Y, Sun Z, Zhang X. 2019. A markerless automatic deformable registration framework for augmented reality navigation of laparoscopy partial nephrectomy. International Journal of Computer Assisted Radiology and Surgery:1–10.