Landmark-free Automatic Digital Twin Registration in Robot-assisted Partial Nephrectomy using a Generic End-to-end Model

Kilian Chandelon[1,2*], Alice Pitout[3], Mathieu Souchaud[2], Julie Desternes[2], Gaëlle Margue[3], Julien Peyras[2], Nicolas Bourdel[4,2], Jean-Christophe Bernhard[3], Adrien Bartoli[5,1,2]

[1]EnCoV, Institut Pascal, UMR6602 CNRS, UCA, Clermont-Ferrand University Hospital, France.
[2]SURGAR - Surgical Augmented Reality, Clermont-Ferrand, France.
[3]Department of Urology, Bordeaux University Hospital, France.
[4]Department of Obstetrics and Gynaecology, Clermont-Ferrand University Hospital, France.
[5]Department of Clinical Research and Innovation, Clermont-Ferrand University Hospital, France.

*Corresponding author(s). E-mail(s): kilian.chandelon@gmail.com;

Abstract

Purpose. Augmented Reality in Minimally-Invasive Surgery has made tremendous progress in organs including the liver and the uterus. The core problem of Augmented Reality is registration, where a preoperative patient's geometric digital twin must be aligned with the image of the surgical camera. The case of the kidney is yet unresolved, owing to the absence of anatomical landmarks visible in both the patient's digital twin and the surgical images. **Methods.** We propose a landmark-free approach to registration, which is particularly well-adapted to the kidney. The approach involves a generic kidney model and an end-to-end neural network, which we train with a proposed dataset to regress the registration directly from a surgical RGB image. **Results.** Experimental evaluation across four clinical cases demonstrate strong concordance with expert-labelled registration, despite anatomical and motion variability. The proposed method achieved an average tumour contour alignment error of $\mathbf{7.3 \pm 4.1}$ mm in $\mathbf{9.4 \pm 0.2}$ ms. **Conclusion.** This landmark-free registration approach meets the accuracy, speed and resource constraints required in clinical practice, making it a promising tool for Augmented Reality-assisted Partial Nephrectomy.

Keywords: Registration; Deep learning; Augmented reality; Digital Twin; Robot-assisted partial nephrectomy

1 Introduction

Minimally-Invasive Surgery has many advantages, including reduced hospital stay, intraoperative blood loss and operative time, but the localisation of the organs' internal structures is challenging. Computer-Assisted Surgery (CAS) addresses such localisation challenges, with a promising solution overlaying the organ's geometric digital twin directly on the surgical video. This Augmented Reality (AR) approach relies on the ability to reconstruct a 3D model of the organ's outer shape and internal structures preoperatively and to register it to the surgical camera video stream intraoperatively. AR is particularly desired in Partial Nephrectomy (PN), which 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). The challenges are to localise the tumours, especially the endophytic ones hidden within the parenchyma, the blood vessels and the urinary excretory tracts. The gold standard uses intraoperative ultrasound (IOUS). Despite its effectiveness, IOUS presents significant limitations, including the images being 2D, the presence of artefacts, the limited depth, the long learning curve and the high dependency on operator experience.

Advanced imaging modalities, which are typically Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), are used preoperatively for diagnosis and treatment planning. They are used routinely to characterise the size, location and morphological complexity of renal tumours. More recently, these imaging data have been used to reconstruct 3D patient twins, providing a highly detailed and personalised representation by means of 3D surface mesh models for the parenchyma and the internal structures. The use of 3D patient twins has recently made its way to the intraoperative setting via Virtual Reality (VR) applications. They however do not resolve the localisation challenges.

AR goes a step further by registering and fusing the 3D patient twin with the intraoperative video. AR for CAS has been researched for a few decades. Technically, AR typically involves three main steps. The process begins with 1) camera calibration, which provides the camera's intrinsic parameters (and the extrinsic parameters for a stereo camera) and establishes the projection function, mapping 3D points expressed in the camera coordinate frame to 2D image points. Next, 2) a registration phase is carried out, where the transformation between the patient's 3D digital twin and the camera's 3D coordinate frames is computed. Finally, 3) a tracking mechanism can be added to maintain dynamic adaptability to organ and camera motion.

Registration is the core step of AR, for which manual, semi-automatic and automatic approaches have been developed for soft organs. The manual approach involves an operator to adjust the orientation, position and scale of the 3D digital twin in the overlay. The semi-automatic and automatic approaches leverage anatomical landmarks visible both on the 3D digital twin and within the intraoperative images, such as vessel bifurcations and organ boundaries, to constrain the registration. The difference is that the semi-automatic approach requires an operator to mark the landmarks in the image, while the automatic approach automatically detects them. Unfortunately however, the kidney presents a significant challenge: its retroperitoneal anatomical location combined with surrounding adipose tissue restrict visualisation, obscure the anatomical primitives, thus preventing the use of methods developed for other organs. During PN robust anatomical primitives defined as landmarks that enable reliable 3D/2D matching are frequently absent. Additionally, these primitives are often procedure-based and vary based on the surgical technique. In rare instances, identifiable structures like the renal hilum or surface irregularities may be visible. The primary visible landmarks are the parenchymal surface and exophytic tumour contours. However, these primitives alone would be insufficient to fully resolve the 3D/2D registration. Neighbouring organs such as the liver, peritoneum and gallbladder have an extent of mobility independent from the kidney, strongly limiting their utility.

The specific accuracy requirements for kidney surgery are not yet formalised to the extent observed in other surgical domains. However, relevant benchmarks can be inferred from comparable procedures. For instance, in hepatectomy, an oncologic margin of 1 cm is mandated for posterior liver tumours that cannot be reliably localised intraoperatively, effectively defining a clinical AR accuracy requirement of approximately 1 cm [1, 2]. Similarly, in myomectomy, myomas smaller than 1 cm often evade detection and, if missed, may necessitate revision surgery [2, 3]. By analogy, in RAPN, achieving sub-centimetric registration accuracy is clinically significant, as it facilitates the precise localisation and resection of small, localised renal masses under 4 cm—cases for which RAPN is the recommended treatment modality according to the European Association of Urology guidelines [4].

We propose a landmark-free automatic registration method applicable to the kidney, as shown in figure 1, to enable AR. We take a neural approach, where the surgical camera image is fed into an end-to-end neural network regressing the desired registration. We thus rid ourselves of the anatomical landmarks. Our method is based on a generic kidney model called the shape prototype, which allows us to train our registration network in a patient-generic manner. A key challenge lies in obtaining training data, as the ground truth (GT) is not directly measurable in this context. To address this limitation, a proxy is used, relying on expert-labelled registrations performed offline through retrospective analysis. We have developed an advanced data labelling system, which allows the experts to visualise both upstream and downstream frames to refine the labelling and ensure accurate training data for the learning-based model.

The technical contributions of this work are fourfold: 1) the development of a landmark-free 3D/2D registration approach based on a shape prototype model enabling an end-to-end solution, 2) a methodology for generating labelled image datasets with shape prototype poses addressing the challenge of limited training data availability, 3) the introduction of a geometry-based automatic kidney canonical frame detection method, and 4) the design and training of an end-to-end inference network for pose estimation ensuring robust and automated performance in clinical scenarios.

The proposed method is validated through quantitative error analysis and AR performance evaluation, assessing the transfer accuracy of the shape prototype pose to the 3D patient twin. The evaluation focuses on enhancing the representation of patient-specific elements including tumours, vessels and urinary excretory tracts to ensure accurate and clinically-relevant alignment.

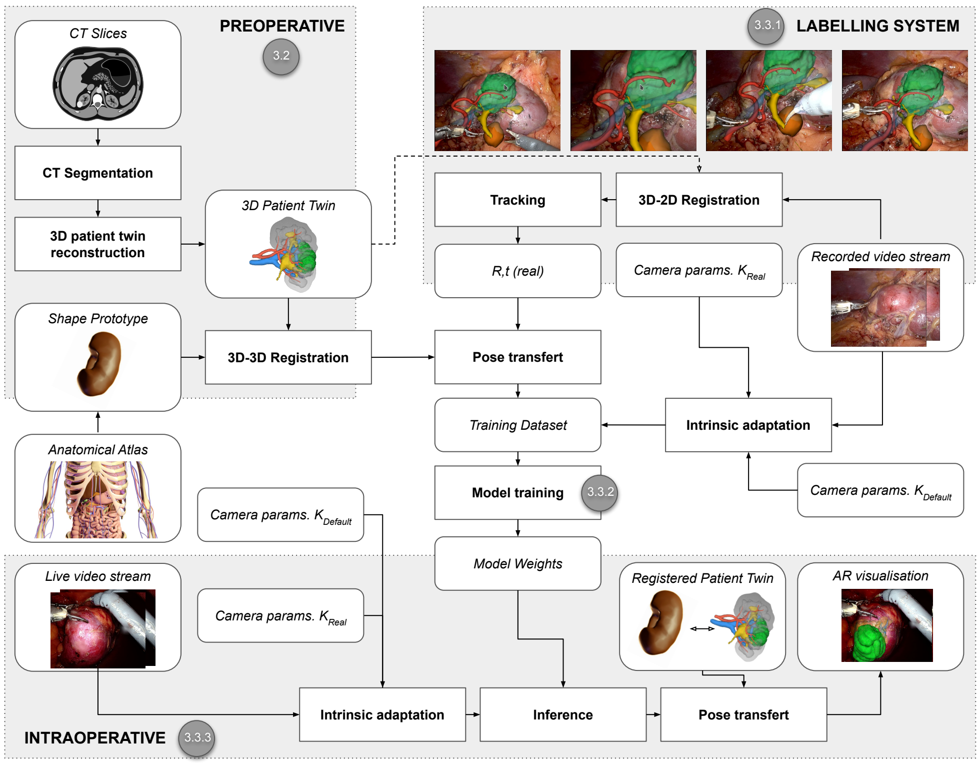

Fig. 1 Pipeline of the proposed landmark-free automatic registration method illustrating the preoperative, intraoperative and labelling steps. The section numbers are given in the gray circles.

2 Related Work

We review related work in registration for other organs of the abdomino-pelvic cavity, then specifically for the kidney, and finally in the problem of human pose estimation, which shares similarities.

2.1 Registration in Abdomino-pelvic Soft-organ Surgery

Aside from the kidney, 3D/2D registration of soft organs in the abdomino-pelvic cavity is an active research field mainly focused on the uterus, liver, pancreas and prostate.

The 3D patient twins are obtained preoperatively from MRI or CT images while the surgical images are captured from the camera hand-held or integrated into a robotic-assistance system.

For the uterus [5] and liver [6–9], early solutions relied on a combination of anatomical landmarks and the organ boundaries to guide the numerical optimisation of the registration parameters. Indeed, these two organs are very well visible and offer landmarks such as the fallopian tube junctions for the uterus and the lower ridge for the liver. More recently, the numerical optimisation has been substituted by a neural network, called Liver Mesh Recovery network, mapping the image landmarks to the desired registration [10].

When landmarks are not clearly visible, alternative strategies have been developed. For the pancreas, registration methods include the manual alignment of anatomical points using an infrared-guided pointing tool [11] and the use of a QR code virtually embedded in the digital twin and manually placed on the organ [12]. For the prostate, registration is achieved by integrating the catheter into the digital twin and using it as a landmark, through its detection using a convolutional neural network (CNN) [13].

2.2 Registration in Partial Nephrectomy

As shown in the surveys [14, 15], registration in PN predominantly relies on manual approaches. Rigid registration typically involves operator-dependent overlay of the 3D patient twin onto selected surgical images [16–24]. Manual methods are also used in radical nephrectomy and thrombectomy procedures [25]. Efforts to enhance 3D/2D registration accuracy have explored non-rigid manual adjustments [22], surface-based [26, 27] and point-cloud based methods [28]. However, these approaches are limited by their operational complexity and lack of scalability in clinical settings. It was attempted to adapt the AR system [5] designed for the uterus to the kidney [29] with limited success, owing to the absence of visible kidney's anatomical features and the absence of a distinct organ silhouette, required for effective registration. The recent CNN-based method automatically computes a rigid registration in two steps [30]. First, the kidney is segmented in the surgical image, its 3D position and scale are computed by fitting an ellipse on the segmentation mask. Second, the orientation is regressed by a 'RotationCNN'. This work has marked differences with ours. First, the RotationCNN is trained from synthetic data obtained by simulating images from the patient twin. It is thus patient-specific and must be trained anew for each patient. Second, the method requires multiple images with smooth camera displacement, quoting "[...] RotationCNN showed a lack of robustness. [...] When it was tested on organs, it was necessary to add to the methodology an alternative third approach, based on Optical Flow (OF), to mitigate the lack of robustness of the RotationCNN." Third, the method does not seem to work, quoting "[...] when the organ presents a more complex structure, such as the kidney [...], the initialisation [of registration] should be manually performed." This is not surprising as training on synthetic images where the organ colour is not available induces a very strong domain gap with the real images.

Therefore, despite extensive research over the past decade, the development of a clinically viable registration solution for PN remains elusive.

2.3 Human Pose Recovery from Monocular Images

Human pose recovery uses a single image to estimate a physically plausible mesh that aligns with the input image and ensures compatibility with the anatomical constraints. It thus has marked similarities to the registration problem at hand. Most contemporary methods rely on combining a parametric body shape model with joint pose representations, such as SCAPE [31] and SMPL [32]. To solve this problem, two main categories have evolved: optimisation-based and learning-based methods [33].

Optimisation-based methods initially estimate a coarse body pose and shape and refine them iteratively by optimising shape parameters to fit multiple image cues such as silhouettes and landmarks [34]. Although effective, these methods are computationally intensive and sensitive to initialisation.

Learning-based methods have seen significant progress following the development of end-to-end frameworks, namely Human Mesh Recovery (HMR) [35], SPIN [36] and the recent PyMAF [37] and PyMAF-X [38]. HMR begins with a ResNet-50 encoder to extract image features which are concatenated with pose, shape and camera parameters and fed to a regression network. This model leverages Iterative Error Feedback (IEF) to iteratively refine predictions. Additionally, a discriminator enhances shape and pose plausibility, especially in scenarios lacking 3D ground truth. Expanding upon foundation models, PyMAF integrates a feature pyramid structure and alignment feedback mechanism for multi-scale refinement. PyMAF-X further extends these capabilities to full-body recovery achieving state-of-the-art results in challenging scenarios. These landmark-free methods are inspiring for PN but adapting them is complex for two main reasons: first, there is no universally-agreed renal shape model, and second, the data are scarce and difficult to collect.

3 Methodology

3.1 Overview and Data Collection

Following the dominant trend in AR, we assume that the virtual information is contained in a patient twin, which is defined as a specific digital twin. While the digital twin is a virtual representation of a physical entity, such as an organ, used for simulation, planning and real-time analysis in medical applications, the patient twin is a specific digital twin that represents an individual patient's unique anatomy, tailored for personalised medical applications. The concept of digital twins may also, in the existing literature, include various types of information such as biological organ characteristics; in this work, we focus on the 3D shape information of the kidney and its internal structures. We use *patient twin* to refer to the preoperative 3D model reconstructed from CT or MRI and containing 3D mesh surfaces representing the kidney's parenchymal shape and the internal structures of interest for AR. We assume that the patient twin is available preoperatively, as in previous work. In addition, we use *shape prototype* to refer to a generic 3D model of the kidney's parenchyma shape, which enables the proposed approach to be patient-generic.

The proposed registration method works in two steps. In step 1), the patient twin is registered to the shape prototype. This is done preoperatively. In step 2), the shape prototype is registered to the camera image. This is done intraoperatively. By combining the transformations obtained from steps 1) and 2), we obtain the desired transformation from the patient twin to the surgical image, required to achieve AR. In our experiments, the proposed registration method is combined with a preoperative patient twin reconstruction method and an intraoperative kidney tracking method, achieving a highly functional system.

We achieve step 2) by an end-to-end neural network mapping the image to the registration. This NN is trained in a supervised manner. We thus have to make crucial choices: the surgical phases which we address and the desired type of transformation. PN is typically divided into 4 to 16 phases, depending on the authors and the level of detail. For our approach, we have used five key phases [39, 40]: kidney exposure, vascular dissection, tumour identification and delineation, tumour resection and tumour bed reconstruction. We focus on the first three phases, occurring before the parenchymal incision and where AR is typically required.

We collected a comprehensive dataset for kidney pose estimation using seven high-resolution videos (1280 × 1024 pixels, 60 Hz) recorded from RAPN procedures performed using the Intuitive Da Vinci Xi surgical system. The corresponding 3D patient twins were reconstructed via contrast-enhanced CT segmentation by the clinical team involved in the surgery. This team included a primary senior surgeon over 8 years of expertise in the field, assisted by a surgeon with at least 2 years of experience in RAPN. The segmentation encompassed five anatomical structures: kidney parenchyma, renal vein, renal artery, urinary tract and tumour. Empirical observations from over 30 cases indicate that rigid local registration adequately satisfies clinical requirements for tumour localisation during PN. While accounting for deformations may offer advantages, especially for extra-organ vessel visualisation.

3.2 Preparation of Preoperative Patient Data

The kidney shape prototype, bean-shaped with a convex lateral and concave medial border, measures 11 cm in length, 6 cm in width and 3 cm in thickness. Although modelled for the right kidney, it is also suitable for the left one, as both are nearly identical in size, differing mainly in anatomical position. It was extracted from the Z-Anatomy digital atlas.

Preoperative registration computes the transformation T_{pt2sp} between the patient twin's parenchyma (pt) and the shape prototype (sp) in three coarse-to-fine steps (figure 2). It begins with 1) canonical frame initialisation, where the patient twin's frame is set at its centre of mass. The Z-axis is aligned to the superior pole based on the eigenvectors of the inertia tensor, while the X-axis is aligned through the hilum. The hilum position is identified by filtering concave vertices near the centre of mass, considering curvature, proximity, anterior direction and normal alignment to refine the candidate points. After establishing the canonical frame, a coarse alignment is performed by aligning the centres of mass and frame axes. Finally, 2) a fine alignment is achieved using an Iterative Closest Point (ICP) algorithm applied forward and backward, followed by 3) a Non-Rigid ICP (NR-ICP) step. This comprehensive process ensures an accurate match between the kidney in the patient twin and the shape prototype.

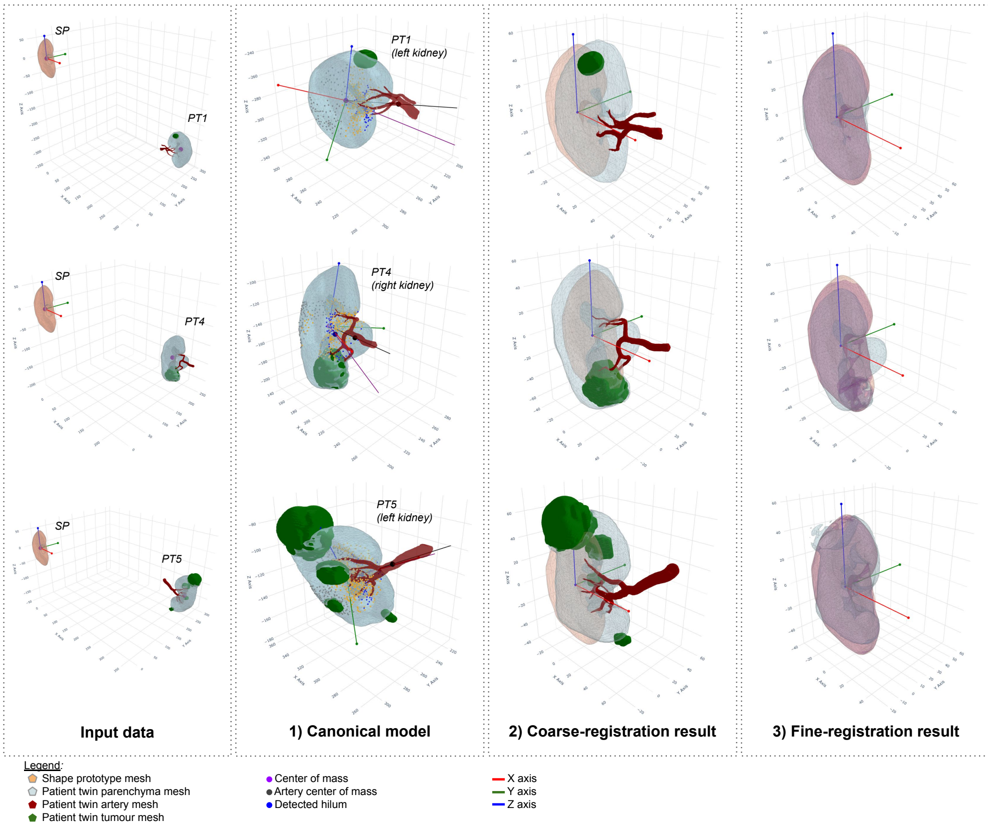

Fig. 2 Coarse-to-fine 3D registration of the Patient Twin (PT) onto the Shape Prototype (SP) in three steps: 1) canonical model initialisation, 2) coarse alignment via the canonical frame and 3) fine-tuning with ICP.

3.3 End-to-end Generic Pose Recovery

We described the proposed method to infer the pose of the shape prototype from an intraoperative image.

3.3.1 Data Labelling

A team of urological surgeons from a high-volume expert center (University Hospital of Bordeaux, France), hereafter referred to as *experts*, routinely using intraoperative ultrasound (IOUS) and virtual reality for image-guided partial nephrectomy, retrospectively annotated the calibrated intraoperative videos. These annotations were conducted by a surgeon with 2 years of experience in RAPN, operating under the supervision of a senior surgeon possessing more than 8 years of experience in the field. A dedicated offline application for manually aligning the 3D patient twin onto keyframes where AR could support the procedure was used. This process was not time-constrained, enabling flexible navigation both upstream and downstream within the video sequences.

Although real-time intraoperative annotation was not feasible, retrospective labelling was carried out by the same surgical team that conducted the procedures, leveraging their in-depth knowledge of patient-specific anatomy. This approach improved consistency, as annotators could select keyframes where anatomical structures were clearly visible and verify information by reviewing temporally adjacent frames. Furthermore, tumour boundaries had been intraoperatively marked with bipolar energy under IOUS guidance (a routine clinical practice) providing anatomical reference points that enhanced annotation accuracy across augmented frames.

Following manual alignment, a rigid feature-based tracking algorithm similar to the method proposed by [5, 41] was subsequently used to temporally and spatially interpolate the position of the 3D patient twin. This yielded seven augmented sequences with expert-verified 3D poses. From 228 manually annotated keyframes representing approximately 40 working hours, 299,659 tracked frames were automatically generated, significantly improving efficiency.

A semi-automatic review ensured positional and orientational consistency across sequences. Automated alerts flagged anomalies such as abrupt frame-to-frame differences or values outside the defined 3D workspace. This approach improved the reliability of labelled data and minimised tracking errors.

The manual expert labels are standardised for deep model training through two steps: alignment of 3D patient twins with the kidney shape prototype from the Z-Anatomy atlas (see section 3.2) and the camera intrinsics are standardised with frame warping (section 3.3.3) to ensure consistency across procedures.

3.3.2 Neural Architecture and Training

We proposed a 3D/2D registration model using a pretrained DINO-BASE-V2 ImageNet backbone as an image encoder, generating feature maps of dimensions $(16, 16, 768)$ from standardised 224×224 RGB endoscopic images. Each image is paired with a pose $P = (R, T)$ between the shape prototype and the camera, with the rotation $R \in SO(3)$ and the translation $T \in \mathbb{R}^3$. These elements form a $(1, 12)$ output tensor. They were shown to be stable, continuous and learnable [42].

The feature maps are concatenated with the current pose parameters and fed into a regression module with three fully connected layers using LeakyReLU activation and dropout to improve gradient flow and prevent overfitting. Pose parameters are iteratively refined over three iterations using an Iterative Error Feedback (IEF) mechanism, aligning the shape prototype with the input image.

The training loss $\mathcal{L} = \lambda\mathcal{L}_{\text{pose}} + (1-\lambda)\mathcal{L}_{\text{mesh}}$ has two terms, mixed by a hyperparameter $\lambda \in [0, 1]$. The 3D mesh loss $\mathcal{L}_{\text{mesh}}$ is defined as the Mean Absolute Error (MAE) of the euclidean distances between the predicted and manually-labelled mesh vertices. The pose loss $\mathcal{L}_{\text{pose}}$ is defined as the MAE between the normalised predicted and manually-labelled 12 pose coefficients.

The model is trained over 100 epochs with a batch size of 16 using the AdamW optimiser—a stochastic optimisation method that decouples weight decay from the gradient update and a learning rate scheduler starting at 0.00001. A curriculum learning strategy gradually increases task complexity, reflecting surgical progression as the organ undergoes transformations like mobilisation and incisions, deviating from its original shape and appearance. Data augmentation is used, including blur, noise and compression of the input images.

3.3.3 Intraoperative Pose Inference

The endoscopic camera is calibrated when surgery starts, determining its intrinsic parameters K_{real}. During model training, these parameters were unavailable and default intrinsics K_{default} were used. To align the intrinsic spaces between the trained model and the calibrated setup, a 2D affine transformation $A = K_{\text{default}}K_{\text{real}}^{-1}$ is used [43]. The model predicts the pose P_{sp} in the shape prototype coordinate frame. The corresponding patient twin pose P_{pt} is obtained by applying the inverse transformation $T_{\text{sp2pt}} = T_{\text{pt2sp}}^{-1}$, yielding $P_{\text{pt}} = P_{\text{sp}}T_{\text{pt2sp}}$.

4 Experimental Results

We validate our method on the collected seven-patient dataset. We use a leave-one-patient-out (LOPO) approach, training the model on six patients' data and testing on the remaining patient.

4.1 Performance Evaluation

The predicted pose, expressed in the patient twin coordinate frame, is compared to the manually labelled pose. Two types of errors were computed: the pose error which includes translation and rotation components and the Mean Absolute Error (MAE) between the mesh vertices. The proposed model was trained on an Ubuntu PC with Intel Core™ i7-13620H and an Nvidia RTX 4070 GPU card.

Table 1 Overall performance evaluation of the proposed method compared to the manual expert labels.

Test patient ID	PN side	Num. frames for eval.	Avg. translation error (mm)	Avg. rotation error (°)	Avg. vertex error (mm)
1	Left	38 267	13.21	0.54	14.56
2	Left	35 077	12.63	0.62	14.02
3	Right	29 449	16.17	0.45	16.84
4	Right	15 952	11.50	0.43	12.32
5	Left	66 577	14.36	0.94	15.93
6	Left	60 237	13.44	0.72	15.14
7	Left	54 100	12.16	0.32	12.67
		MEAN	13.35	0.57	14.50
		STD	1.55	0.21	1.65

The results in table 1 show the proposed method's accuracy. Low mean translation and rotation errors, alongside minimal MAE of the vertices, indicate high concordance with manual expert labels. Despite anatomical and motion variability, the framework achieved a consistently strong performance. This underscores its capability to autonomously align the 3D patient twin with the surgical camera image, making it a valuable asset for AR applications.

4.2 Application to Surgical Guidance

The proposed method is integrated into a broader framework designed to augment still images and image sequences during RAPN. The intraoperative view is augmented with hidden anatomical structures transferred from the preoperative 3D patient twin, using either (1) real-time inference for continuous AR video or (2) automatic registration of an initial frame followed by rigid tracking [5, 41].

We have conducted a preliminary retrospective evaluation of the clinical relevance of our system, focusing on time efficiency and accuracy, specifically targeting the registration error for the tumour. To facilitate quantitative evaluation, we selected 4 cases with exophytic tumours, which allowed direct error measurement. The error was measured by comparing the projected tumour contour, predicted by registration, with the real observed tumour contour, as shown in figure 3. The proposed AR system allows the user to customise the rendering transparency level for each type of structure. As a reference, manual alignment errors ranged from 3 mm to 9 mm without time constraints, increasing to 20 mm when limited to 10 seconds. In our experiments, we observed that registration becomes significantly more difficult for third-party operators who is not the treating surgeon and has no prior context.

Our method achieves an average execution time of 9.4 ± 0.2 ms (106 FPS) using PyTorch 2.5.1, encompassing pre-processing (i.e., input standardisation and GPU transfer) and post-processing (i.e., output transfer to patient twin). The average tumour contour alignment error was 7.3 ± 4.1 mm. To our knowledge, this is the only viable fully automatic 3D/2D registration approach currently available for RAPN. For comparison, state-of-the-art methods in liver surgery report average errors of 30 mm to 36 mm for patient-specific and generic models respectively [44]. Our approach delivers significantly lower error without manual input or patient-specific modelling, which is highly promising for its clinical reliability and practical deployment.

5 Conclusion

We have proposed a landmark-free automatic registration method designed for application in partial nephrectomy, directly regressing the pose parameters from an image. The proposed approach is well-suited to the temporal and human resource constraints typical of clinical practice. Today, it represents the only feasible solution capable of achieving a registration accuracy of 7.3 ± 4.1 mm (measured on 4 patients) without reliance on the expertise of the operating surgeon. Nonetheless, given the early-stage feasibility evaluation, and although consistency with expert annotations has been observed, claims of clinical reliability remain preliminary and require validation against surgical outcomes. The current experiments are based on a small cohort (n=7), with

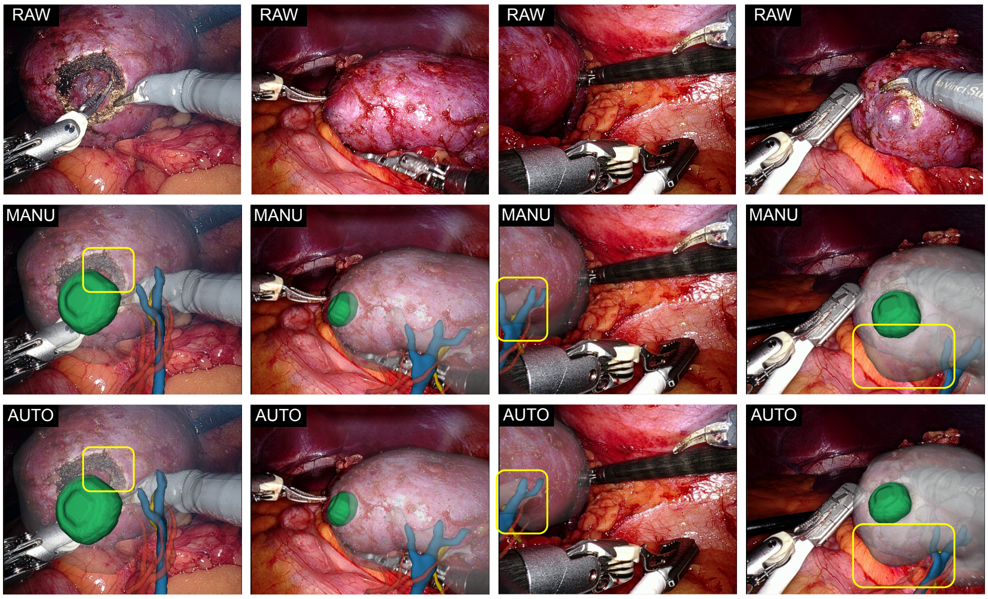

Fig. 3 Augmented Reality overlays for a representative test case using the proposed method. All images correspond to the tumour identification and delineation phase. The top row shows the endoscopic view, the second row displays the manually labelled 3D patient twin pose (annotation) and the third row shows the predicted 3D patient twin pose generated by our method.

tumour error analysis limited to four cases, which restricts statistical power and generalisability. To mitigate overfitting and support initial generalisation, we employed a leave-one-patient-out (LOPO) cross-validation strategy. Our current method addresses only rigid registration but remains relevant for surgical guidance within the restricted kidney volume surrounding the tumour, where rigid assumptions are locally valid. However limited exposure remains a limitation of the current system and is not yet fully characterised for retroperitoneal route due to a lack of data.

Future work will focus on 1) expanding the dataset to include retrospectively collected cases (n=30) to improve generalisability, 2) integrating a deformation component into the model to account for organ shape variations, particularly in vascular structures, and 3) broadening its applicability to a wider range of surgical phases. Based on this technical 3D/2D automatic registration module, clinical features such as tumour infiltration depth, proximity to the collecting system, surgical margins and perfusion territories will be explored to further support surgical decision-making. In the short term, this method is intended to serve as a complementary tool to IOUS with the potential to overcome its current limitations over time.

Acknowledgments

We express our gratitude to UroCCR: French Kidney Cancer Research Network for supplying the anonymised surgical videos used in this study. This work was supported by the French Government through the France 2030 program (ANR-21-RHUS-0015).

Declarations

Conflict of Interest: The authors declare that they have no conflict of interest. **Ethical approval:** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors. **Informed consent:** Informed consent was obtained from the patients included in the study.

References

[1] Zhong, F.-P., Zhang, Y.-J., Liu, Y., Zou, S.-B.: Prognostic impact of surgical margin in patients with hepatocellular carcinoma: a meta-analysis. Medicine **96**(37), 8043 (2017)

[2] Khojasteh, S.B., Fuentes-Jimenez, D., Pizarro, D., Espinel, Y., Bartoli, A.: Misnerf: Neural radiance fields in minimally-invasive surgery. International Journal of Computer Assisted Radiology and Surgery (2025)

[3] Desai, P., Patel, P.: Fibroids, infertility and laparoscopic myomectomy. Journal of gynecological endoscopy and surgery **2**(1), 36 (2011)

[4] Bex, A., Albiges, L., Bedke, J., Capitanio, U., Dabestani, S., Hora, M., Klatte, T., Kuusk, T., Lund, L., Marconi, L., *et al.*: Renal cell carcinoma. In: European Association of Urology Guidelines (2025)

[5] Collins, T., Pizarro, D., Gasparini, S., Bourdel, N., Chauvet, P., Canis, M., Calvet, L., Bartoli, A.: Augmented reality guided laparoscopic surgery of the uterus. IEEE Transactions on Medical Imaging **40**(1), 371–380 (2020)

[6] Espinel, Y., Calvet, L., Botros, K., Buc, E., Tilmant, C., Bartoli, A.: Using multiple images and contours for deformable 3d-2d registration of a preoperative ct in laparoscopic liver surgery. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24, pp. 657–666 (2021). Springer

[7] Haouchine, N., Roy, F., Untereiner, L., Cotin, S.: Using contours as boundary conditions for elastic registration during minimally invasive hepatic surgery. In: 2016

IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 495–500 (2016). IEEE

[8] Robu, M.R., Ramalhinho, J., Thompson, S., Gurusamy, K., Davidson, B., Hawkes, D., Stoyanov, D., Clarkson, M.J.: Global rigid registration of ct to video in laparoscopic liver surgery. International Journal of Computer Assisted Radiology and Surgery **13**, 947–956 (2018)

[9] Modrzejewski, R., Collins, T., Seeliger, B., Bartoli, A., Hostettler, A., Marescaux, J.: 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 (2019)

[10] Labrunie, M., Pizarro, D., Tilmant, C., Bartoli, A.: Automatic 3d/2d deformable registration in minimally invasive liver resection using a mesh recovery network. In: MIDL, pp. 1104–1123 (2023)

[11] Okamoto, T., Onda, S., Yasuda, J., Yanaga, K., Suzuki, N., Hattori, A.: Navigation surgery using an augmented reality for pancreatectomy. Digestive surgery **32**(2), 117–123 (2015)

[12] Tang, R., Yang, W., Hou, Y., Yu, L., Wu, G., Tong, X., Yan, J., Lu, Q.: Augmented reality-assisted pancreaticoduodenectomy with superior mesenteric vein resection and reconstruction. Gastroenterology research and practice **2021**(1), 9621323 (2021)

[13] Checcucci, E., Piana, A., Volpi, G., Piazzolla, P., Amparore, D., De Cillis, S., Piramide, F., Gatti, C., Stura, I., Bollito, E., Massa, F., Di Dio, M., Fiori, C., Porpiglia, F.: Three-dimensional automatic artificial intelligence driven augmented-reality selective biopsy during nerve-sparing robot-assisted radical prostatectomy: A feasibility and accuracy study. Asian Journal of Urology **10**(4), 407–415 (2023)

[14] Khaddad, A., Bernhard, J.-C., Margue, G., Michiels, C., Ricard, S., Chandelon, K., Bladou, F., Bourdel, N., Bartoli, A.: A survey of augmented reality methods to guide minimally invasive partial nephrectomy. World journal of urology **41**(2), 335–343 (2023)

[15] Della Corte, M., Quarà, A., De Cillis, S., Volpi, G., Amparore, D., Piramide, F., Piana, A., Sica, M., Di Dio, M., Alba, S., Porpiglia, F., Checcucci, E., Fiori, C.: 3d virtual models and augmented reality for radical prostatectomy: a narrative review. Chinese Clinical Oncology **13**(4), 56–56 (2024)

[16] Su, L.-M., Vagvolgyi, B.P., Agarwal, R., Reiley, C.E., Taylor, R.H., Hager, G.D.: Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3d-ct to stereoscopic video registration. Urology **73**(4), 896–900

(2009)

[17] Nakamura, K., Naya, Y., Zenbutsu, S., Araki, K., Cho, S., Ohta, S., Nihei, N., Suzuki, H., Ichikawa, T., Igarashi, T.: Surgical navigation using three-dimensional computed tomography images fused intraoperatively with live video. Journal of endourology **24**(4), 521–524 (2010)

[18] Pratt, P., Mayer, E., Vale, J., Cohen, D., Edwards, E., Darzi, A., Yang, G.-Z.: An effective visualisation and registration system for image-guided robotic partial nephrectomy. Journal of Robotic Surgery **6**, 23–31 (2012)

[19] Chen, Y., Li, H., Wu, D., Bi, K., Liu, C.: Surgical planning and manual image fusion based on 3d model facilitate laparoscopic partial nephrectomy for intrarenal tumors. World journal of urology **32**, 1493–1499 (2014)

[20] Wang, D., Zhang, B., Yuan, X., Zhang, X., Liu, C.: 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 (2015)

[21] Porpiglia, F., Fiori, C., Checcucci, E., Amparore, D., Bertolo, R.: 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 (2018)

[22] Porpiglia, F., Checcucci, E., Amparore, D., Piramide, F., Volpi, G., Granato, S., Verri, P., Manfredi, M., Bellin, A., Piazzolla, P., Autorino, R., Morra, I., Fiori, C., Mottrie, A.: Three-dimensional augmented reality robot-assisted partial nephrectomy in case of complex tumours : a new intraoperative tool overcoming the ultrasound guidance. European urology **78**(2), 229–238 (2020)

[23] Schiavina, R., Bianchi, L., Chessa, F., Barbaresi, U., Cercenelli, L., Lodi, S., Gaudiano, C., Bortolani, B., Angiolini, A., Bianchi, F.M., Ercolino, A., Casablanca, C., Molinaroli, E., Porreca, A., Golfieri, R., Diciotti, S., Marcelli, E., Brunocilla, E.: Augmented reality to guide selective clamping and tumor dissection during robot-assisted partial nephrectomy: a preliminary experience. Clinical genitourinary cancer **19**(3), 149–155 (2021)

[24] Amparore, D., Piramide, F., Pecoraro, A., Verri, P., Checcucci, E., De Cillis, S., Piana, A., Busacca, G., Manfredi, M., Fiori, C., Porpiglia, F.: 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 (2022)

[25] Amparore, D., Checcucci, E., Piramide, F., Busacca, G., Volpi, G., De Cillis, S., Sica, M., Verri, P., Piana, A., Di Dio, M., Fiori, C., Porpiglia, F.: Robotic

vena cava thrombectomy with three-dimensional augmented reality guidance. European Urology Open Science **62**, 43–46 (2024)

[26] Altamar, H.O., Ong, R.E., Glisson, C.L., Viprakasit, D.P., Miga, M.I., Herrell, S.D., Galloway, R.L.: Kidney deformation and intraprocedural registration: a study of elements of image-guided kidney surgery. Journal of endourology **25**(3), 511–517 (2011)

[27] Amparore, D., Checcucci, E., Piazzolla, P., Piramide, F., De Cillis, S., Piana, A., Verri, P., Manfredi, M., Fiori, C., Vezzetti, E., Porpiglia, F.: Indocyanine green drives computer vision based 3d augmented reality robot assisted partial nephrectomy: the beginning of "automatic" overlapping era. Urology **164**, 312–316 (2022)

[28] Zhang, X., Wang, J., Wang, T., Ji, X., Shen, Y., Sun, Z., Zhang, X.: A markerless automatic deformable registration framework for augmented reality navigation of laparoscopy partial nephrectomy. International journal of computer assisted radiology and surgery **14**, 1285–1294 (2019)

[29] Teluob, G., Calvet, L., Espinel, Y., Savareux, L., Guandalino, M., Ravel, A., Bourdel, N., Chauvet, P., Guy, L., Chabrot, P., Bartoli, A.: Preliminary trial of augmented reality performed on a regular and a robot-assisted laparoscopic partial nephrectomies. Videourology **33**(3) (2019)

[30] Piana, A., Amparore, D., Sica, M., Volpi, G., Checcucci, E., Piramide, F., De Cillis, S., Busacca, G., Scarpelli, G., Sidoti, F., Alba, S., Piazzolla, P., Fiori, C., Porpiglia, F., Di Dio, M.: Automatic 3d augmented-reality robot-assisted partial nephrectomy using machine learning: our pioneer experience. Cancers **16**(5), 1047 (2024)

[31] Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: Scape: shape completion and animation of people. ACM SIGGRAPH 2005 Papers, 408–416 (2005)

[32] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: Smpl: A skinned multi-person linear model. Seminal Graphics Papers: Pushing the Boundaries, Volume 2, 851–866 (2023)

[33] Tian, Y., Zhang, H., Liu, Y., Wang, L.: Recovering 3d human mesh from monocular images: A survey. IEEE transactions on pattern analysis and machine intelligence (2023)

[34] Guan, P., Weiss, A., Balan, A.O., Black, M.J.: Estimating human shape and pose from a single image. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 1381–1388 (2009). IEEE

[35] Kanazawa, A., Black, M.J., Jacobs, D.W., Malik, J.: End-to-end recovery of

human shape and pose. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7122–7131 (2018)

[36] Kolotouros, N., Pavlakos, G., Black, M.J., Daniilidis, K.: Learning to reconstruct 3d human pose and shape via model-fitting in the loop. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2252–2261 (2019)

[37] Zhang, H., Tian, Y., Zhou, X., Ouyang, W., Liu, Y., Wang, L., Sun, Z.: Pymaf: 3d human pose and shape regression with pyramidal mesh alignment feedback loop. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 11446–11456 (2021)

[38] Zhang, H., Tian, Y., Zhang, Y., Li, M., An, L., Sun, Z., Liu, Y.: Pymaf-x: Towards well-aligned full-body model regression from monocular images. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(10), 12287–12303 (2023)

[39] Farinha, R., Breda, A., Porter, J., Mottrie, A., Van Cleynenbreugel, B., Vander Sloten, J., Mottaran, A., Gallagher, A.G., RAPN-Delphi Surgeons Group: International expert consensus on metric-based characterization of robot-assisted partial nephrectomy. European Urology Focus **9**(2), 388–395 (2023)

[40] De Backer, P., Peraire Lores, M., Demuynck, M., Piramide, F., Simoens, J., Oosterlinck, T., Bogaert, W., Shan, C.V., Van Regemorter, K., Wastyn, A., Checcucci, E., Debbaut, C., Van Praet, C., Farinha, R., De Groote, R., Gallagher, A., Decaestecker, K., Motttrie, A.: Surgical phase duration in robot-assisted partial nephrectomy: A surgical data science exploration for clinical relevance. Diagnostics **13**(21), 3386 (2023)

[41] Chandelon, K., Bartoli, A.: Tracking better, tracking longer: automatic keyframe selection in model-based laparoscopic augmented reality. International Journal of Computer Assisted Radiology and Surgery **17**(8), 1507–1511 (2022)

[42] Geist, A.R., Frey, J., Zobro, M., Levina, A., Martius, G.: Learning with 3d rotations, a hitchhiker's guide to so (3). arXiv preprint arXiv:2404.11735 (2024)

[43] Fuentes-Jimenez, D., Pizarro, D., Casillas-Pérez, D., Collins, T., Bartoli, A.: Deep shape-from-template: Single-image quasi-isometric deformable registration and reconstruction. Image and Vision Computing **127**, 104531 (2022)

[44] Labrunie, M., Pizarro, D., Tilmant, C., Bartoli, A.: Generic liver modelling with application to mini-invasive surgery guidance. In: International Conference on Medical Imaging and Computer-Aided Diagnosis, pp. 219–229 (2024). Springer