International Journal of Computer Assisted Radiology and Surgery (2021) 16:1181–1188
https://doi.org/10.1007/s11548-021-02419-9

ORIGINAL ARTICLE

Preclinical evaluation of a markerless, real-time, augmented reality guidance system for robot-assisted radical prostatectomy

Megha Kalia[1] · **Apeksha Avinash**[1] · **Nassir Navab**[2] · **Septimiu Salcudean**[1]

Received: 19 March 2021 / Accepted: 21 May 2021 / Published online: 2 June 2021

Abstract

Purpose Intra-operative augmented reality (AR) during surgery can mitigate incomplete cancer removal by overlaying the anatomical boundaries extracted from medical imaging data onto the camera image. In this paper, we present the first such completely markerless AR guidance system for robot-assisted laparoscopic radical prostatectomy (RALRP) that transforms medical data from transrectal ultrasound (TRUS) to endoscope camera image. Moreover, we reduce the total number of transformations by combining the hand–eye and camera calibrations in a single step.

Methods Our proposed solution requires two transformations: TRUS to robot, ${}^{\mathrm{DV}}T_{\mathrm{TRUS}}$, and camera projection matrix, $\mathbf{M}$ (i.e., the transformation from endoscope to camera image frame). ${}^{\mathrm{DV}}T_{\mathrm{TRUS}}$ is estimated by the method proposed in Mohareri et al. (in J Urol 193(1):302–312, 2015). $\mathbf{M}$ is estimated by selecting corresponding 3D-2D data points in the endoscope and the image coordinate frame, respectively, by using a CAD model of the surgical instrument and a preoperative camera intrinsic matrix with an assumption of a projective camera. The parameters are estimated using Levenberg–Marquardt algorithm. Overall mean re-projection errors (MRE) are reported using simulated and real data using a water bath. We show that $\mathbf{M}$ can be re-estimated if the focus is changed during surgery.

Results Using simulated data, we received an overall MRE in the range of 11.69–13.32 pixels for monoscopic and stereo left and right cameras. For the water bath experiment, the overall MRE is in the range of 26.04–30.59 pixels for monoscopic and stereo cameras. The overall system error from TRUS to camera world frame is 4.05 mm. Details of the procedure are given in supplementary material.

Conclusion We demonstrate a markerless AR guidance system for RALRP that does not need calibration markers and thus has the capability to re-estimate the camera projection matrix if it changes during surgery, e.g., due to a focus change.

Keywords Surgical augmented reality · Robot-assisted radical prostatectomy · Markerless camera projection matrix estimation

Introduction

Robot-assisted laparoscopic radical prostatectomy (RALRP) is a frequently performed cancer surgery during which the entire prostate is removed. Because prostate cancer is not visible with standard endoscopic equipment, the surgeon does not know the cancer boundaries and needs to guess them; thus, in many cases cancer is left behind, leading to possible cancer recurrence, or healthy tissue is removed, leading to decreased quality of life [2,13].

Augmented reality (AR) can mitigate this problem by overlaying the cancerous tissue, segmented from preoperative imaging, such as magnetic resonance imaging (MRI), onto the endoscope camera image. Therefore, in recent years several AR guidance systems for RALRP have been proposed [7,11,14].

In [7], an intra-operative surface reconstruction of the prostate from stereo endoscopic camera images is registered to a mesh generated from MRI. In [11], mesh generated from MRI is manually aligned with the pose of the prostate in the endoscope camera image. The registered image is rendered in a head mounted display (HMD) [7] or into the surgical console [11]. These approaches either utilize a marker to find the

✉ Megha Kalia
mkalia@ece.ubc.ca

[1] Electrical and Computer Engineering, University of British Columbia, 2329 West Mall, Vancouver, BC V6T 1Z4, Canada

[2] Computer Aided Medical Procedures, Technical University of Munich Boltzmannstraße 15, 85748 Garching bei München, Germany

camera parameters for the surface reconstruction step [7] or require an alignment step [11].

In [10,12], ultrasound is used as an intermediate step to register MRI to the patient. First, a transrectal ultrasound (TRUS) system is registered to the da Vinci robot [1]. Then, patient-specific preoperative MRI is deformably registered to the intra-operative TRUS. The da Vinci console shows a rendered image of the prostate surface obtained from the registration process, and the real-time ultrasound as well as a preoperative MRI plane that are pointed at by the surgeon's instrument. The system does not use any external markers during surgery. However, it is only a partial AR system, because these images are not overlaid onto the endoscopic camera view. Instead, they are shown in a separate "TilePro" window, with the surgeon having to infer MRI image location from the rendered instrument tip and TRUS transducer image.

Thus, in prior work, intra-operative calibration must be used for rendering or manual alignment is required; these interfere with the surgical workflow and thus present a hindrance in the adoption of AR [8]. Alternatively, the AR system does not present the data overlaid onto the actual surgical view.

This paper presents an AR guidance system for RALRP that transforms the medical imaging data from the TRUS coordinate system to the camera image coordinate system without using any external calibration markers before or during surgery. Our AR pipeline needs two transformations to be estimated. The first is the transformation from the TRUS to the robot base, performed as in [10]. The second transformation finds the camera projection matrix **M** from the endoscope coordinate frame, also called the endoscope control manipulator (ECM) frame, to the camera image coordinate frame. **M** is estimated from corresponding 3D and 2D data points located in the robot and the camera image coordinate systems, respectively, using the work of Ye *et al.*, 2016, [15] with key changes.

Unlike [15], our method uses a preoperative camera intrinsic matrix and an identity hand–eye transformation to render the CAD model of the surgical instrument and therefore does not require markers during surgery. Then, 3D-2D correspondences are used to jointly estimate the camera and hand–eye transformations, in the form of a camera projection matrix, in a single step.

We performed two evaluation experiments on simulated and real data and report overall re-projection errors, from TRUS to the camera image frame. Through another experiment we also demonstrate that during surgery, if the endoscope is moved or the focus is changed, the projection matrix can be re-estimated by collecting the data points by moving the surgical instrument in the surgical field of view.

The novelty of our proposed AR guidance pipeline consists of combining the hand–eye and camera calibration steps into a single step and evaluating its efficacy. The significance of the approach lies in its simplicity, speed and lack of external markers. The markerless scheme makes it possible to re-estimate the projection matrix if the focus is changed midsurgery or if the region of interest is changed within the camera view. The projection matrix re-estimation can be performed recursively, in less than a minute, allowing for adaptation to the scene during surgery, without the need to stop the surgical flow. Although we propose a guidance system for RALRP, the methods used are suitable for any robotic surgery where instruments can be seen in the TRUS and camera image and where the camera can be modeled as a perspective camera.

Methods

TRUS to DV calibration, ${}^{\mathrm{DV}}T_{\mathrm{TRUS}}$

To find the transformation ${}^{\mathrm{DV}}T_{\mathrm{TRUS}}$ between the TRUS and the da Vinci base coordinate frame, we used the method described in [10]. The surgeon palpates at the gas–tissue boundary of the prostate intra-operatively using the surgical instrument. The surgical instrument tip can be seen in the TRUS image and is manually segmented by clicking on the instrument tip using a computer mouse. The corresponding points in robot base coordinate frame are obtained from the robot's forward kinematics data. Thus, in total four points at the apex, base and two lateral sides of the prostate are collected. The rigid body transformation between the two coordinate frames is estimated using a least-squares method. After applying this transformation, TRUS points can be transformed to the robot coordinate frame by applying the transformation, ${}^{\mathrm{DV}}T_{\mathrm{TRUS}}$. Once the medical data are registered to the robot's base coordinate system, it can easily be transformed to the ECM coordinate frame, by multiplication with the matrix ${}^{\mathrm{ECM}}T_{\mathrm{DV}}$, obtained from the forward kinematics data of the robot.

Camera projection matrix estimation, M

The camera projection matrix represents the hand–eye calibration matrix (i.e., the transformation from the ECM to the camera world coordinate system) and the camera intrinsic matrix (i.e., the transformation from the camera world to the image coordinate system). In many instances, including ours, where the medical imaging data is in the robot's ECM coordinate system, estimating the projection matrix directly from the ECM to the image coordinate space in a single step is fast and sufficient. Therefore, we modeled the camera as a projective camera and jointly estimated the hand–eye and camera calibration in the form of the projection matrix **M**,

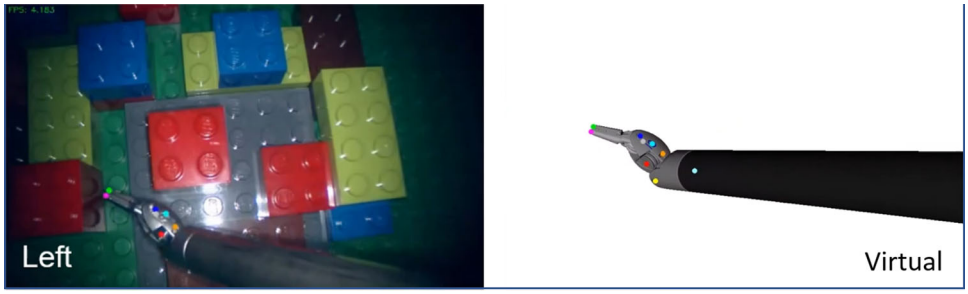

Fig. 1 (Left) Image showing the detected key points (colored) in the real camera image of the surgical instrument. (Right) Image showing the corresponding predefined key points on the CAD model of the surgical instrument in the virtual camera image

where $\mathbf{M} \in R^{3\times4}$. We also included the first two lens distortion parameters, $\mathbf{k}$, of the radial lens distortion model [4] in the camera model as two radial distortion parameters are usually sufficient to model lens distortion [16].

To find $\mathbf{M}$, we located corresponding 3D-2D points in the robot and image coordinates, respectively, using a CAD model by the method proposed by [15]. Fourteen manually determined, fixed key points are located on the CAD model. Unlike [15], the CAD model is rendered in the virtual world using only a preoperative camera intrinsic matrix, an identity hand–eye transformation and robot's forward kinematics data. The preoperative camera intrinsic matrix can be estimated many days before surgery using a checkerboard by the method described in [16]. Note that preoperative camera intrinsic matrix does not correspond to the correct camera intrinsic matrix during surgery as the focus is adjusted after inserting the endoscope inside the patient. However, this matrix is sufficient to render the CAD model on screen. This is the key difference between our method and [15], which makes our proposed method markerless during surgery. Then, quantized gradient orientations (QGO) features are calculated in the bounding boxes defined around the key points in the virtual image and on the entire real image. Then by template matching, key points are located on the real images [Fig. 1]. The CAD model corresponds to links 4 and 5 of the da Vinci manipulator. Thus, the corresponding key points are obtained in the robot base coordinate frame, which then were transformed to the ECM coordinate frame.

The data collection using CAD model can be done easily by moving the surgical instrument at various locations in the surgical field of view within seconds. This is the procedure that we follow for the evaluation experiments described in the subsequent sections.

To find the transformation $\mathbf{M}$, the corresponding 3D-2D pairs $\{\mathbf{x}, \mathbf{u}^l\}$ for the monoscopic case and pairs $\{\mathbf{x}, \mathbf{u}^l, \mathbf{u}^r\}$ for stereoscopic case are considered, where $\mathbf{x}$ are the 3D points in the ECM frame and $\mathbf{u}^l$ and $\mathbf{u}^r$ are the corresponding image points. The notation $\mathbf{x}_i$ and $\mathbf{u}_i$ denote the ith ECM world and ith left or right image points, respectively.

To account for the noise in the data points, 3D data points are also included in the optimization scheme, where the optimized data points are $\hat{\mathbf{x}}$. The following objective function is minimized:

$$\min_{\mathbf{M}_l,\mathbf{M}_r,\hat{\mathbf{x}},k_1^c,k_2^c} \sigma_u^{-1}\mathcal{E}_{\Pi}\left(\mathbf{M}^c, \hat{\mathbf{x}}, k_j^c\right) + \sigma_x^{-1}\mathcal{E}_{\Theta}\left(\hat{\mathbf{x}}\right) \ . \tag{1}$$

where $\mathbf{M}^c \in \mathcal{P}_{\text{Projective}}, \ \ c \in \{l, r\}$, and

$$\mathcal{E}_{\Theta}\left(\hat{\mathbf{x}}\right) = \sum_c \sum_i \left(\hat{\mathbf{x}}_i - \mathbf{x}_i\right)^2, \quad \mathcal{E}_{\Pi}\left(\mathbf{M}^c, \hat{\mathbf{x}}\right) = \sum_c \sum_i \left(\hat{\mathbf{u}}_i^c - \mathbf{M}^c\hat{\mathbf{x}}_i\right)^2,$$

and σ_u and σ_x are the standard deviation of noise in the image and the robot world points, respectively. $\hat{\mathbf{u}}_i^c$ is the distorted image point where the x and y elements of $\hat{\mathbf{u}}_i^c$ are $\hat{u}_i$ and $\hat{v}_i$, respectively, and can be written as follows:

$$\hat{u}_i = u_i + \left(u_i - u_p\right)\left[\sum_m^2 k_m\left(u_{ni}^2 + v_{ni}^2\right)^m\right] \tag{2}$$

$$\hat{v}_i = v_i + \left(v_i - v_p\right)\left[\sum_m^2 k_m\left(u_{ni}^2 + v_{ni}^2\right)^m\right] \tag{3}$$

where v_{ni} and u_{ni} are ith normalized image points and $\mathbf{u}_p^c = \left(u_p^c, v_p^c\right)$ are center of radical distortion [4].

The Levenberg–Marquardt (LM) algorithm is used to jointly optimize the above problem. Since the convergence of iterative algorithms is dependent on initialization, we initialized the LM algorithm with the solution of direct linear transform (DLT) [4]. For better numerical stability, the DLT image points were normalized [5] and outliers were rejected by using the random sample consensus (RANSAC) algorithm.

In summary, the following transformations were applied to transform the real-time TRUS image to the right or left endoscope camera image coordinate system [Fig. 2]:

$${}^{i}T_{\text{TRUS}} = \mathbf{M} \times^{\text{ECM}} T_{\text{DV}} \times^{\text{DV}} T_{\text{TRUS}} \tag{4}$$

where ${}^{\text{DV}}T_{\text{TRUS}}$, ${}^{\text{ECM}}T_{\text{DV}}$ and $\mathbf{M}$ are transformations from TRUS to the da Vinci robot base, robot base to ECM and ECM to camera image (camera projection matrix), respectively.

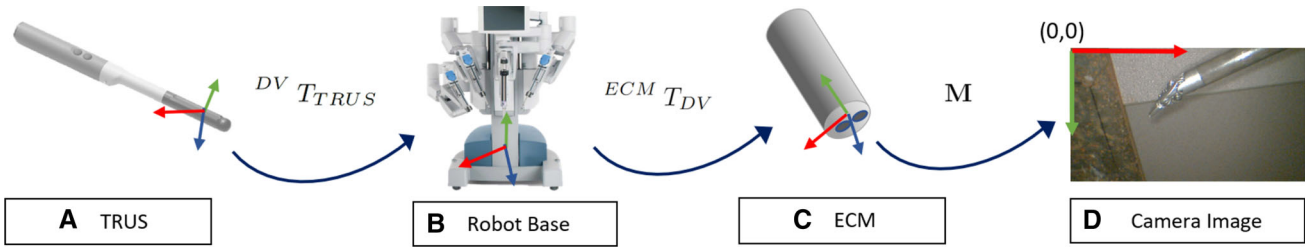

Fig. 2 A schematic of transformations involved in transforming the data from the TRUS to the endoscope camera image coordinate frame

Evaluation and results

To evaluate the overall accuracy and robustness of our methods, we performed experiments with real and simulated data, respectively. All the experiments are done on the da Vinci Si system and the images of resolution 960 × 540 pixels.

Simulation experiment

Simulation experiment was carried out to analyze the robustness of our numerical methods by eliminating other sources of errors from the experiment setup. The primary sources of error in estimating M are the user bias in manually selecting the instrument tip in the TRUS image for ${}^{\mathrm{DV}}T_{\mathrm{TRUS}}$ estimation and the error in the CAD model-based key point detection on the surgical instrument.

For the simulation experiment, we prepared data by performing all proposed calibration steps required to bring the data from the TRUS coordinate system to the camera coordinate frame on the da Vinci robot. The calibration steps were performed on the real setup to obtain transformations in the correct configuration, such that the TRUS points are visible in the camera image. Thus, after estimating the TRUS to the robot base transformation [Sect. 2.1] and camera projection matrix [Sect. 2.2] we had ground-truth transformations ${}^{\mathrm{DV}}T^{g}_{\mathrm{TRUS}}$ and projection matrix $\mathbf{M}^{g}$, respectively.

In total, 500 data points were generated in the TRUS coordinate frame covering a volume of 60 mm × 40 mm × 40 mm. The average prostate size lies well within this volume [9]. These points were transformed to the image coordinate space by applying the ground-truth transformations in the sequence given in Eq. (4). Subsets of these point set were used for the ${}^{\mathrm{DV}}T_{\mathrm{TRUS}}$ calculation and validation, and for reporting overall mean re-projection error from TRUS to image coordinate frame.

To calculate $\mathbf{M}$, 1000 data points were generated in the endoscope coordinate frame. These points were transformed to the image space using the ground truth $\mathbf{M}^{g}$. To simulate realistic settings, random noise of 0.6 mm standard deviation (SD) was added to the data points in the TRUS, robot and endoscope coordinate frames. Similarly, 0.5 pixel SD noise was added to the points in the right and the left images.

To estimate the overall re-projection error, the entire chain of calibrations was repeated 30 times to transform the data points from the TRUS to the left image coordinate space for monocular case and to the left and right camera frames in stereo case. For the ${}^{\mathrm{DV}}T_{\mathrm{TRUS}}$ calculation, at each iteration we randomly sampled 8–10 points in TRUS coordinates and estimated ${}^{\mathrm{DV}}T_{\mathrm{TRUS}}$ using the method described in Sect. 2.1. For $\mathbf{M}$ estimation, at each iteration, we randomly sampled 500 corresponding data points in the endoscope left and right image coordinate frames, and then solved for $\mathbf{M}$ using DLT and LM initialized with DLT solution with two distortion parameters [Sect. 2.2]. We will refer to these methods as "DLT" and "LM" in the rest of the paper. Estimations were done for both the stereoscopic and the monoscopic cases. The overall mean re-projection error is reported on 486–492 ground-truth points in the TRUS coordinate frame, not used in ${}^{\mathrm{DV}}T_{\mathrm{TRUS}}$ estimation.

The average overall re-projection errors for monocular and stereoscopic cases for DLT and LM methods are reported in Table 1.

Evaluation using real data using a water bath

For evaluation on real data, we used the water bath approach from [6]. We prepared a grid of 25 cross-wire points using a nylon wire and custom 3D printed structure. Using the 3D printed structure, the grid could be moved in 1 cm increments. In the water bath, the cross-wire points could be seen in both the endoscope stereo camera images and the TRUS image as shown in Fig. 3. We manually segmented the 3D cross-wire points in the TRUS image. To locate the corresponding 2D cross-wire points in the camera images, we first drained the water out to avoid distortions from refraction, without moving the setup. Then the left and right camera images were captured and 2D cross-wire points were manually segmented. To capture the corresponding set of points, we placed the grid at various configurations and depths, such that it is visible in both TRUS and camera images. Thus, we collected a set of 115 data points. Thus, we had ground truth for 3D TRUS and 2D left and right data points. For evaluation, we first estimated ${}^{\mathrm{DV}}T_{\mathrm{TRUS}}$ and $\mathbf{M}$ following the procedures described in Sect. 2. We also evaluated $\mathbf{M}$ at this

Table 1 Overall mean re-projection error from TRUS to image coordinate frame on simulated data

	Monoscopic	Stereoscopic	
	Left	Left	Right
DLT	11.80 ± 3.62 pixels	11.68 ± 3.51 pixels	11.69 ± 3.44 pixels
LM	11.88 ± 3.78 pixels	13.31 ± 4.39 pixels	13.32 ± 4.46 pixels

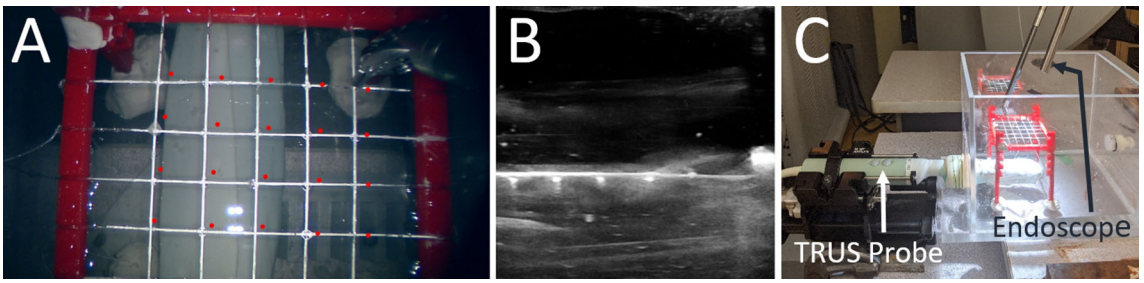

Fig. 3 Sample endoscope camera and TRUS images from water bath experiment. **a** Image showing cross-wire points in the camera image (white) and re-projected corresponding TRUS points (red), after applying transformations ${}^{DV}T_{TRUS}$ and **M**. **b** Sagittal image showing cross-wire points in the TRUS image. **c** Image showing the water bath experiment set up

step to analyze the effect of estimation errors in **M** on the overall mean re-projection error. We divided data collected for **M** estimation into 70% and 30% for model estimation and validation, respectively.

We obtained a camera mean re-projection error of 9.42 pixels and 10.8 pixels using DLT and LM, respectively. For the stereo case, for the left camera, the mean re-projection errors of 9.31 pixels and 12.19 pixels and for the right camera, mean re-projection errors of 9.76 pixels and 11.84 pixels were obtained using DLT and LM, respectively. To estimate the overall mean re-projection error, the ground-truth TRUS cross-wire points were transformed to the camera image by applying the sequence of transformations given in Eq. (4). The visual results are shown in Fig. 3. The overall mean re-projection errors for the monoscopic case using DLT and LM are 27.89 pixels and 26.04 pixels, respectively. For the stereo left camera, the overall mean re-projection errors, using DLT and LM, were 27.80 pixels and 30.59 pixels, respectively. Similarly, for the right stereo camera, the overall mean re-projection errors of 29.81 pixels and 26.20 pixels were obtained using DLT and LM, respectively. Re-projected TRUS cross-wire points are shown in Fig. 3a. The error in the camera world frame was estimated to be 4.05 mm. For details of the procedure, refer to supplementary material.

Evaluation of the camera projection matrix retrieval after a focus change

To demonstrate the feasibility of re-estimating the projection matrix intra-operatively and to assess it, we conducted an experiment. We moved the endoscope at three different depths simulating the scenario of moving the endoscope while performing surgery. At each depth position, we placed the surgical instrument and changed the focus of the endoscope using the pedal at the da Vinci console, such that the instrument is in focus. Then to determine the projection matrix, we moved the surgical instrument at different locations in the camera's field of view in approximately the same depth plane. Using the method described in Sect. 2.2, we collected the corresponding data points in image and robot coordinates and estimated the projection matrix. To demonstrate that at each depth after focus change the projection matrix has converged to a correct solution, we used two validation schemes. The aim of the first scheme is to quantify error in **M** estimation at each depth level, while the aim of the second scheme is to visually asses the **M** estimation (Fig. 4).

In the first scheme, we used 70% of the data, collected using the CAD model, for projection matrix estimation and used the remaining 30% for validation for calculating the mean re-projection error, similar to previous evaluation experiments. In the second scheme, at each depth level we placed a checkerboard, filling most of the camera image, and touched the corners of the checkerboard manually using the surgical instrument. A checkerboard with a larger square size was used at the greater depth to fill a significant area of the image. Thus, we found the checkerboard corner points in robot coordinate frame. For determining the corresponding checkerboard corner points in the image coordinate frame, we used the algorithm given in [3]. On both datasets, we reported errors for monoscopic and stereoscopic conditions. The results are given in Table 2.

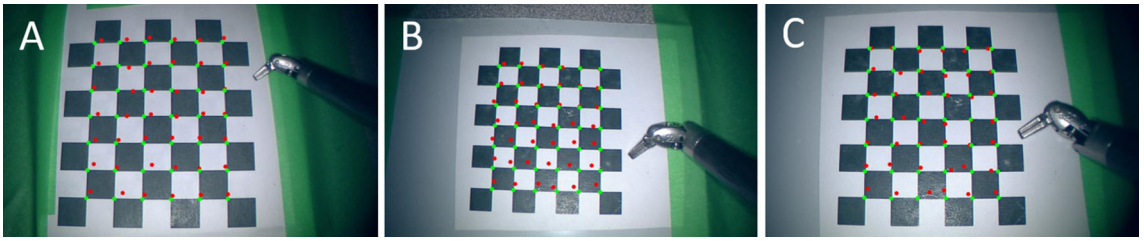

Fig. 4 Images showing re-projected data points after applying projection matrices estimated at three different depth levels. Depths from the endoscope are decreasing from (**a**–**c**). The endoscope was moved at each depth level and the focus was adjusted, in a single session. In the given image, the size of the visible surgical instrument can be referred to for inferring the distance from the endoscope. Different checkerboards are used in images (**a**) and (**b**) and (**c**). Green and red are ground-truth and re-projected corresponding robot points, respectively

Table 2 Re-projection errors for projection matrix estimation at three depth levels after focus change (the values are given in pixel units)

	DLT			LM		
	Monoscopic	Stereoscopic		Monoscopic	Stereoscopic	
	Left	Left	Right	Left	Left	Right
Focus 1	6.89	7.22	7.82	7.52	7.47	8.72
Focus 2	10.02	10.77	11.50	9.89	10.39	10.15
Focus 3	7.85	7.67	8.20	7.40	7.37	8.55

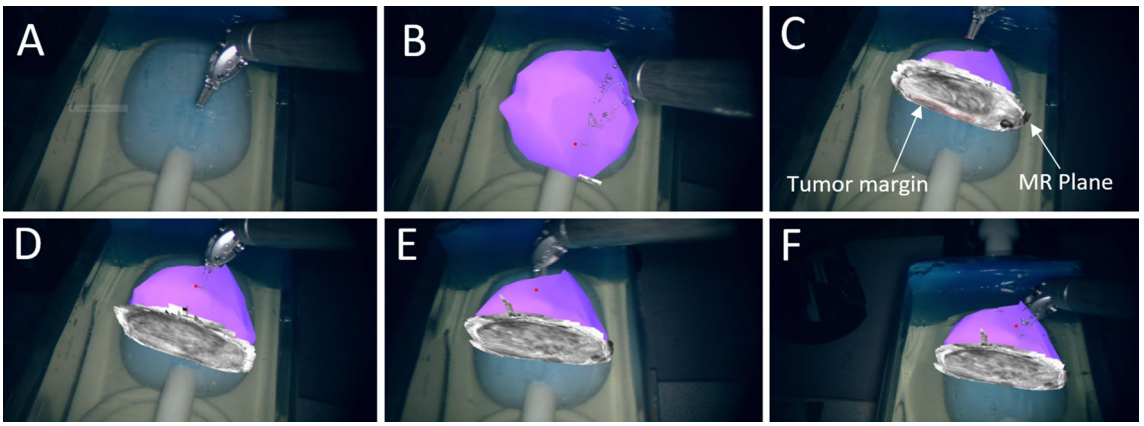

Fig. 5 Images showing our AR guidance system with overlaid finite element mesh (FEM) generated from a patient-specific preoperative MRI which is deformably registered to the TRUS image of the prostate phantom. Note that the TRUS image of the phantom is not shown in the visualization. **a** Prostate phantom (blue) without the AR visualization. **b** Prostate phantom with overlaid FEM mesh generated from MRI. **c** Image showing the control of transverse MR plane using surgical instrument. Tumors could be seen in red in the MR image. **d**–**f** Images demonstrating the tracked AR visualization when the position and orientation of the endoscope camera changes

Discussion and conclusion

We presented an AR guidance system for RALRP that transforms real-time TRUS imaging data to the camera image space without depending on external calibration marker before or during surgery; only the surgical instrument visible in the surgical field of view is utilized to estimate ${}^{\mathrm{DV}}T_{\mathrm{TRUS}}$ and **M**, without changing the surgical pipeline significantly. Moreover, our system combines the camera and the hand–eye calibration steps into one, thus simplifying the AR pipeline.

Through simulation and experiments, we showed that the our system provides promising visual results when rendered on the camera image [Figs. 3 and 5]. We also presented an exemplary real-time AR guidance system using a prostate phantom where a patient-specific preoperative MRI is deformably registered to the real-time TRUS image of the prostate phantom. The transverse MR plane could be controlled in real time using the surgical instrument to see the delineated tumors rendered at the correct location.

We obtained an overall system error of 4.05 mm which is comparable to the error of 4.56 mm reported in [6]. Comparable errors, despite significantly different hand–eye and camera calibration methods, could be due to an error in ${}^{\mathrm{DV}}T_{\mathrm{TRUS}}$ estimation step. Thus, the observation warrants a

future study to identify the source of error in the current calibration pipeline.

Unlike in previously proposed AR guidance systems for RALRP [7,11], we also reported errors in pixels. This is important for AR surgical applications where the data is overlaid onto the endoscope camera image, as the re-projection errors in pixels give a clear picture of the expected visual offsets between the rendered medical image and the endoscope image of the anatomy. For many surgical applications, the offsets between the two modalities could be the bottleneck.

The overall mean re-projection error, from TRUS to camera image, on simulation and real data is $\approx$ 12 pixels and $\approx$ 27 pixels, respectively. Since the prostate usually fills a significant image region in the endoscope camera image during the nerve sparing procedure in RALRP, our visualization, even with the given visual offsets, can become a valuable tool for surgical decisions. For example, when the tumor lies at the base of the prostate, a surgeon can preserve more tissue surrounding the urinary sphincter to maintain continence. If the tumor is on the medial region, more of the NVB can be spared, helping to maintain sexual function. Our system keeps the surgeon informed of the cancer location during surgery, without requiring difficult mental mapping between different coordinate frames.

The mean re-projection errors are larger in the water bath experiment when using real data. Most of the error is likely due to manually segmenting the cross-wire points in the TRUS image, which is difficult because they span multiple TRUS planes. In the future, we plan to use 3D images and automatic feature detection. Furthermore, reflections at the junction of the two nylon wires (as shown in Fig. 3b) make the selection of true cross-wire points more difficult. As reported in Sect. 3.2, the camera projection matrix estimation step has a camera mean re-projection error of $\approx$ 10 pixels. Small errors in roll angle due to erroneous point selection when combined with the errors in the camera projection matrix could lead to larger offsets in the endoscope image as the images are generally magnified. Moreover, the endoscope is usually in a titled position during surgery; thus, the offsets are perceived to be larger than their true size due to perspective.

We also demonstrated through an experiment that using our methods $\mathbf{M}$ can be re-estimated if the focus has changed during surgery. The re-projected data points when rendered on the checkerboard image give higher than expected visual error, given the low value of re-projection errors [Fig. 5]. This could be because of the errors arising from the user's depth perception while manually selecting the checkerboard corners and kinematics errors at the patient side manipulator (PSM) which are within a couple of millimeters from the PSM base. Interestingly we found that errors in $\mathbf{M}$ estimation are lower when the surgical instrument is moved in a specific depth plane without traversing other depth planes, for data collection, using the CAD model. More experiments will be done in future to look into ways to develop protocols to lower the camera re-projection error.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11548-021-02419-9.

Acknowledgements We are thankful for financial support provided by the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council (NSERC) and the Charles Laszlo Chair in Biomedical Engineering held by Professor Salcudean. Canada Foundation of Innovation (CFI) provided infrastructure support. We would also like to thank Intuitive Surgical for providing the da Vinci Research API and support.

Declarations

Conflicts of interest The authors have no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent This article does not contain patient data.

References

1. Adebar TK, Yip MC, Salcudean SE, Rohling RN, Nguan CY, Goldenberg SL (2012) Registration of 3d ultrasound through an air-tissue boundary. IEEE Trans Med Imaging 31(11):2133–2142
2. Davis JW, Kreaden US, Gabbert J, Thomas R (2014) Learning curve assessment of robot-assisted radical prostatectomy compared with open-surgery controls from the premier perspective database. J Endourol 28(5):560–566
3. Geiger A, Moosmann F, Car Ö, Schuster B (2012) Automatic camera and range sensor calibration using a single shot. In: 2012 IEEE international conference on robotics and automation, pp 3936–3943. IEEE
4. Hartley R, Zisserman A (2003) Multiple view geometry in computer vision. Cambridge University Press, Cambridge
5. Hartley RI (1997) In defense of the eight-point algorithm. IEEE Trans Pattern Anal Mach Intell 19(6):580–593
6. Kalia M, Mathur P, Tsang K, Black P, Navab N, Salcudean S (2020) Evaluation of a marker-less, intra-operative, augmented reality guidance system for robot-assisted laparoscopic radical prostatectomy. Int J Comput Assist Radiol Surg 15:1225–1233
7. Kolagunda A, Sorensen S, Mehralivand S, Saponaro P, Treible W, Turkbey B, Pinto P, Choyke P, Kambhamettu C (2018) A mixed reality guidance system for robot assisted laparoscopic radical prostatectomy. In: OR 2.0 context-aware operating theaters, computer assisted robotic endoscopy, clinical image-based procedures, and skin image analysis, pp 164–174. Springer
8. Linte CA, Davenport KP, Cleary K, Peters C, Vosburgh KG, Navab N, Jannin P, Peters TM, Holmes DR III, Robb RA (2013) On mixed reality environments for minimally invasive therapy guidance: systems architecture, successes and challenges in their implementation from laboratory to clinic. Comput Med Imaging Graph 37(2):83–97

9. Mitterberger M, Horninger W, Aigner F, Pinggera GM, Steppan I, Rehder P, Frauscher F (2010) Ultrasound of the prostate. Cancer Imaging 10(1):40
10. Mohareri O, Ischia J, Black PC, Schneider C, Lobo J, Goldenberg L, Salcudean SE (2015) Intraoperative registered transrectal ultrasound guidance for robot-assisted laparoscopic radical prostatectomy. J Urol 193(1):302–312
11. Porpiglia F, Fiori C, Checcucci E, Amparore D, Bertolo R (2018) Augmented reality robot-assisted radical prostatectomy: preliminary experience. Urology 115:184
12. Samei G, Tsang K, Kesch C, Lobo J, Hor S, Mohareri O, Chang S, Goldenberg SL, Black PC, Salcudean S (2019) A partial augmented reality system with live ultrasound and registered preoperative MRI for guiding robot-assisted radical prostatectomy. Med Image Anal. https://doi.org/10.1016/j.media.2019.101588
13. Silberstein JL, Eastham JA (2014) Significance and management of positive surgical margins at the time of radical prostatectomy. Indian J Urol: IJU: J Urol Soc India 30(4):423
14. Thompson S, Penney G, Billia M, Challacombe B, Hawkes D, Dasgupta P (2013) Design and evaluation of an image-guidance system for robot-assisted radical prostatectomy. BJU Int 111(7):1081–1090
15. Ye M, Zhang L, Giannarou S, Yang GZ (2016) Real-time 3d tracking of articulated tools for robotic surgery. In: International conference on medical image computing and computer-assisted intervention, pp 386–394. Springer
16. Zhang Z (2000) A flexible new technique for camera calibration. IEEE Trans Pattern Anal Mach Intell 22(11):1330–1334