Augmented Reality Guided Laparoscopic Surgery of the Uterus

Toby Collins, Daniel Pizarro, Simone Gasparini, Nicolas Bourdel, Pauline Chauvet, Michel Canis, Lilian Calvet, Adrien Bartoli

Augmented Reality Guided Laparoscopic Surgery of the Uterus

T. Collins, D. Pizarro, S. Gasparini, N. Bourdel, P. Chauvet, M. Canis, L. Calvet and A. Bartoli

***Abstract*—A major research area in Computer Assisted Intervention (CAI) is to aid laparoscopic surgery teams with Augmented Reality (AR) guidance. This involves registering data from other modalities such as MR and fusing it with the laparoscopic video in real-time, to reveal the location of hidden critical structures. We present the first system for AR guided laparoscopic surgery of the uterus. This works with pre-operative MR or CT data and monocular laparoscopes, without requiring any additional interventional hardware such as optical trackers. We present novel and robust solutions to two main sub-problems: the initial registration, which is solved using a short exploratory video, and update registration, which is solved with real-time tracking-by-detection. These problems are challenging for the uterus because it is a weakly-textured, highly mobile organ that moves independently of surrounding structures. In the broader context, our system is the first that has successfully performed markerless real-time registration and AR of a mobile human organ with monocular laparoscopes in the OR.**

***Index Terms*—Augmented Reality, Laparoscopy, Gynecology, Registration, Tracking, Markerless, Surgical Navigation**

I. Introduction

A laparoscopic surgeon consults pre-operative images such as MR or CT to localize hidden structures such as tumors and major vessels. However, it can be difficult, even for experienced surgeons, to accurately predict their positions during surgery. One of the main goals of CAI is to ease this task by enriching laparoscopic images with data from pre-operative MR or CT using AR [1], [2]. The key technical challenge is non-rigid registration of soft-body organs. Once achieved, the position of the hidden structures can be augmented onto the laparoscopic video. A major open challenge is achieving registration accurately, reliably and in real-time.

We present the first complete AR pipeline for the uterus using a segmented pre-operative 3D model and monocular laparoscopic images, with novel technical contributions at various stages. This facilitates important clinical applications, including AR-assisted resection of lesions such as uterine fibroids. We specifically target monocular laparoscopes because their use is far more widespread than stereo laparoscopes in standard (non-robotic) laparoscopic procedures. This is because of several factors including cost, setup time, image resolution, port size and display comfort. However, the technical challenges are much greater with monocular laparoscopes.

D. Pizarro (dani.pizarro@gmail.com) is with the Department of Electronics, Alcala University, Madrid, Spain. S. Gasparini, N. Bourdel, L. Calvet and A. Bartoli ({lilian.calvet,adrien.bartoli}@gmail.com) are with EnCoV, IP, UMR 6602 CNRS, Université Clermont Auvergne, Clermont-Ferrand, France. S. Gasparini (simone.gasparini@irit.fr) is also with the University of Toulouse, Toulouse INP – IRIT, France. N. Bourdel, P. Chauvet, M. Canis ({nbourdel,pchauvet,mcanis}@chu-clermontferrand.fr) are with the Department of Gynecological Surgery, Clermont-Ferrand, France. T. Collins (toby.collins@ircad.fr) completed work on this project entirely while at EnCoV, before moving to IRCAD, Strasbourg, France.

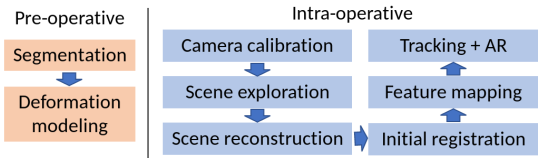

Fig. 1: Main phases of our AR pipeline

The main phases of our AR pipeline are illustrated in Fig. 1 and explained in detail in §3. Pre-operatively, the uterus is segmented from an MR image and its deformation properties are modelled. Intra-operatively, first, automatic *camera calibration* is performed to determine the laparoscope's intrinsics such as its focal length, by withdrawing the scope from the patient and viewing a hand-held calibration planar target (OpenCV). Next is *scene exploration*, where the laparoscope is re-inserted and the uterus surface is viewed by movement of the laparoscope and uterus. Next is *scene reconstruction*, where the uterus surface is reconstructed in 3D using dense multi-view-stereo (MVS). Next is *initial registration*, where a 3D registration is performed to align the uterus model to the interventional reconstruction. Next is *feature mapping*, where texture, in the form of keypoints such as SIFT [3] or SURF [4], is associated to the model's surface. Finally, is *tracking and AR*, where the model is tracked in real-time using robust keypoint matching with the laparoscope live video, and the registered model is visualized to the surgeon via AR. Our main contributions concentrate on the challenging problems of initial registration, feature mapping and tracking.

II. Related Work

A. Scope

The types of AR-guided laparosurgery with the most clinical impact involve the fusion of *pre-operative* medical image data from MR or CT, to which we therefore limit the scope of this section. A broader perspective, including fusion with intra-operative images such as Cone Beam CT (CBCT), is given in [5]. An important categorization of approaches is whether they work with monocular [6] or stereo laparoscopes [7]–[12]. As ours is in the former category, we mostly focus on that category. Any approach that works with monocular scopes can be applied to stereo scopes. The converse is not true, because existing methods using stereo laparoscopes require depth maps obtained by stereo triangulation [13]. Recently, Convolutional Neural Networks (CNNs) have been trained

to recover depth from endoscopic images (colonoscopy [14] and bronchoscopy [15]), however they are yet unsuccessful with laparoscopy because of the large variability in image content. The availability of depthmaps fundamentally changes a registration problem, because they provide 3D-to-3D registration constraints. This contrasts monocular registration, where 3D-to-3D constraints are unavailable. Like us, previous approaches solve monocular registration in two stages: an initial registration stage and a tracking stage.

B. Initial Registration

Despite considerable research, there exists no automatic and robust solution to the initial registration with a soft-body organ. State-of-the-art approaches tend to be organ-specific and have mainly focused on registering the liver [11], [16]–[19], the prostate [9] and the kidney [6], [12], [20]. So far the only existing approaches for the initial registration with monocular images require a manual registration [6] and an interactive Graphical User Interface (GUI), which is not practical in real OR conditions. Of the stereo methods, some perform registration with a manual GUI [9], [11] and others perform it semi-automatically with manually located landmarks [7], [8], [10], [12] or using 3D surface features [19]. In some works the registration is refined by Iterative Closest Point (ICP) [10], [12]. The initial registration requires non-visual constraints to prevent unlikely or physically implausible deformations. Various models have been used, including rigidity [12], deformation smoothness with 3D splines [8] and bio-mechanics [10], [11], [19]. There is no general consensus on the best model to use, as it depends heavily on available boundary conditions, available knowledge of mechanical tissue properties, and computational resources. Recent works have built on deep learning [21]–[23], showing promising results but they do not work with real monocular images.

A general limitation of the previous works is that they only use one monocular or one stereo image pair to constrain the initial registration. This is limiting because registration accuracy depends strongly on how much organ surface is visible. This can be very small, particularly for larger organs such as the liver and uterus, leading to poor registration.

C. Tracking

Almost all monocular approaches rely on the detection and tracking of features, either artificial fiducial markers [9] inserted on the organ, or natural keypoints. The former are invasive and generally not practical. The latter are sensitive to illumination changes, large camera motion and occlusion. These factors critically affect the performance of tracking as they restrain the capability of maintaining the registration, and hence AR visualization, for long periods of time, especially if the organ is deformed or occluded by, *e.g.*, the surgery tools.

To date, only one previous work has been capable of robust long duration tracking of the kidney (several minutes) without artificial fiducials [6]. This work has however two main limitations. Firstly, only one reference image is used, which means features only exist on the surface region visible in the reference image. Tracking therefore breaks down if the organ is seen from strong viewpoint changes. This is a common situation for the uterus, because unlike the kidney it is highly mobile, and is often moved by the surgeon's assistant with a cannula. Secondly, the initial registration is performed manually, which is not practical in real OR conditions. In our approach, we overcome both of these limitations.

Markerless tracking is also addressed by visual Simultaneous Localisation and Mapping (SLAM) [24], [25]. In SLAM, a 3D representation or *map* of the environment is incrementally built and updated and, at the same time, the camera is localized w.r.t. the map. However, SLAM in laparoscopy is not yet reliable enough for routine clinical use. This is because monocular SLAM systems assume a rigid scene and hence are incapable of tracking a mobile organ such as the uterus, as we show in §IV-C. A deformable SLAM method has been recently introduced [26]. The principle is promising but the method requires the scene to be a single deforming object.

D. Contributions

This work describes the first complete pipeline to provide AR-guided uterine laparoscopic surgery without artificial markers or tracking equipment. It is therefore strongly compatible with existing workflows and hospital equipment. To achieve this we present technical innovations that overcome the limits of previous works.

This paper is a distillation and extension of contributions from three workshop papers [27]–[29]. Four significant landmark results have been achieved for the first time in laparoscopic AR in this work, which we now summarize. Firstly, we show that dense *in-vivo* 3D reconstruction is achievable with a monocular laparoscope with Structure-from-Motion (SfM) and MVS. Following this, SfM and MVS have been applied by other groups in related problems. Reconstruction is provided up to an unknown scale factor. This ambiguity is always present in motion-based methods, including SfM and SLAM. Secondly, we recover the reconstruction's absolute scale, by solving it simultaneously with the initial registration via numerical optimization. Thirdly, we show how the organ's silhouette (contours in the laparoscopic images corresponding to the organ's boundary) can be used to significantly improve the initial registration. This fruitfully complements information from the scene reconstruction (Fig. 2), and is particularly important for organs that are very smooth and lack strong geometric details, such as the uterus. Without contours, the model can incorrectly slide over the reconstruction and drift from the correct solution. Fourthly, we track a mobile organ (one that, like the uterus, can move independently of surrounding structures) using robust keypoint matching within tracking-by-detection. This allows the organ to be tracked over long durations (several dozens of minutes) in the presence of difficulties including partial views, occlusions and when the organ moves out of the laparoscope's field-of-view.

This paper presents several improvements of our complete AR pipeline towards clinical use and to reduce manual effort during the initial registration. Specifically, three main contributions considerably improved our workshop papers. We (*i*) ease contour marking by using a touchscreen requiring only non-precise finger strokes, thus making the approach much faster

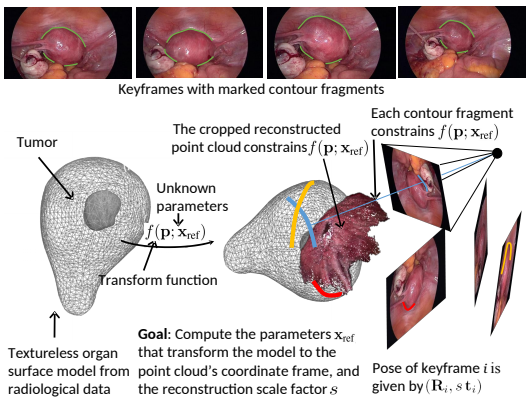

Fig. 2: The initial registration problem illustrated on a patient case. Four keyframes from the exploratory video are shown in the first row with their associated silhouette contour fragments.

and more practical in the OR. We (*ii*) propose a semi-automatic mechanism to collect the frames for 3D reconstruction of the organ by SfM. The tool guides the surgeon in the exploratory phase to acquire sharp images that are also sufficiently spread in space. We show that this has a dramatic impact on the reconstruction quality and tracking robustness. We (*iii*) considerably improve the tracking algorithm using a robust real-time SIFT detector, which increases tracking robustness and smoothness. We present a manual mechanism to update the model's keypoints over time to overcome small appearance and geometric changes in the organ during tracking. This allows us to handle gradual texture changes of the organ during surgery and significantly improves tracking quality.

The result is a system that we have tested 12 times live in the OR, which we recall has not been previously achieved for any organ, including the uterus, using monocular laparoscopes and a markerless approach.

III. Methodology

A. Pre-operative Data Requirements

The system requires a segmented pre-operative 3D organ model, which comprises the organ's surface mesh, and meshes of internal structures to be visualized with AR. For the uterus, internal structures are typically the cavity, tumors and safe-tissue margins. Our approach does not require a specific organ deformation model to be used, because to date there is no clear consensus on the best one to use for registering organs.

We require two interfaces to the deformation model. The first is the *transform function* $f(\mathbf{p};\mathbf{x}_t) : \Omega \to \mathbb{R}^3$. Given the model's parameters $\mathbf{x}_t$ at time t, it transforms a 3D point $\mathbf{p}$ of the model's 3D domain $\Omega \subset \mathbb{R}^3$ to the laparoscope coordinate frame. The second interface is the internal energy function $E_{\text{int}}(\mathbf{x}_t) : \mathbb{R}^d \to \mathbb{R}^+$. This returns the internal energy for transforming the organ according to $\mathbf{x}_t$ (for mechanical models, the strain energy induced by soft-tissue deformation), used to regularize the deformation. We only require that f and E_{int} be continuously differentiable, which is satisfied by virtually all models of interest.

B. Registration Pipeline Overview

Our task is registration: to compute $\mathbf{x}_t$ for a given live monocular laparoscopic image. We break it down into the *initial registration stage* and the *tracking stage*. The initial registration is important for two main reasons. Firstly, the different patient posture between pre-operative and intra-operative steps and the insufflation may induce a deformation of the organ. Initial registration estimates this change of shape from the pre-operative to the intra-operative state, or *reference state*. Secondly, during the tracking stage, we assume that the organ does not undergo significant deformations so that the tracking stage can be reasonably modeled with rigid motion. The initial registration allows us to associate texture with the organ's surface, necessary to achieve tracking.

Formally, the two stages of registration break down $f(\mathbf{p};\mathbf{x}_t)$ as $f(\mathbf{p};\mathbf{x}_t) = M(f(\mathbf{p};\mathbf{x}_{\text{ref}});\mathtt{R}_t,\mathbf{t}_t)$. Here $\mathbf{x}_{\text{ref}}$ denotes the organ's unknown deformation for the reference state. The function $M(\cdot;\mathtt{R}_t,\mathbf{t}_t) : \mathbb{R}^3 \to \mathbb{R}^3$ denotes the unknown update transform at time t, parameterized by a 3D rotation $\mathtt{R}_t \in \mathcal{SO}_3$ and translation $\mathbf{t}_t \in \mathbb{R}^3$.

In practice, the rigidity assumption during tracking is reasonable because during live AR guidance the surgeon does not significantly deform the organ. We emphasize that the intended use of AR is to assist spatial comprehension of internal structures and intra-operative resection planning. Typically this is done by guiding the marking of a tumor resection plane on the uterus surface with a coagulation instrument. Such marking is standard practice in uterine surgery. During the actual resection, where strong deformation and topological changes occur, AR visualization is deactivated because the surgeon follows the coagulation marks.

To provide real-time AR, only the tracking stage needs to be real-time. To minimize workflow interruption, we require the initial registration to be computed in no longer than a few minutes. The manual pre-processing takes on average $2\,\text{min}$ and the optimisation $1\,\text{min}$. Tracking is an optimized implementation in C++/CUDA and runs at $\sim 25\,\text{fps}$.

C. Solving the Initial Registration

1) Solution Overview: Fig. 2 shows the initial registration problem, which is challenging to solve for two main issues: *(i)* the model to be registered is textureless and *(ii)* the registration is non-rigid. To overcome these problems, our approach includes a dense 3D reconstruction of the organ's surface using an *exploratory video* and SfM/MVS reconstruction, for which mature and open-source methods exist [30]. Given this reconstruction, we solve the initial registration with numerical optimization of a system that combines data constraints (organ-to-reconstruction distances and contour fragment distances) with constraints from the model's internal energy.

2) Interventional 3D Reconstruction: During the exploratory video, the uterus is rotated by the surgeon's assistant using the cannula (Fig. 3). It moves independently of background structures, so the scene cannot be reconstructed using SLAM, which requires the scene to be globally rigid. We developed a new tool to capture sharp keyframes with significant mutual spatial displacement. We extract SIFT keypoints from

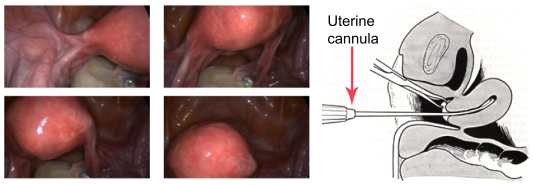

Fig. 3: Example keyframes from an exploratory video of a uterus undergoing cannula motion.

a frame and track them along the frames of the live feed [31]. When the average 2D keypoint displacement exceeds a threshold (default $100\,\text{pixels}$) or too many keypoints are lost w.r.t. the initial frame (default $35\,\%$), we consider that the camera has sufficiently moved. We are then ready to acquire a new keyframe and the UI asks the surgeon to hold the camera still. We then extract and track new keypoints and the keyframe is acquired when the maximum displacement remains below a threshold (default $3\,\text{pixels}$) for the last $M = 15$ frames. Then the process starts again until at least $N = 15$ keyframes are acquired. We use a touch-screen interface to manually segment the uterus roughly in the keyframes, so that the background is masked and only the organ is reconstructed. This takes a few seconds per keyframe. We then run a state-of-the-art SfM/MVS open source library [32]. As output we obtain a dense 3D point cloud $\mathcal{Q} \stackrel{\text{def}}{=} \{\mathbf{q}_j\}$, $\mathbf{q}_j \in \mathbb{R}^3$, and, for each keyframe, the relevant camera pose matrix $\mathtt{M}_i \in \mathcal{SE}_4$, holding the rotation matrix $\mathtt{R}_i \in \mathcal{SO}_3$ and translation vector $\mathbf{t}_i \in \mathbb{R}^3$ w.r.t. the point cloud. We recall that all MVS reconstructions are only computed up to an unknown global scale factor s. We solve for s jointly with registration in §III-E.

3) Silhouette Contours: An occluding contour is a boundary in a 2D image where a surface in the foreground occludes a surface behind it (Fig. 2). There are two types: *self-occluding contours*, which are formed where the object self-occludes, and *silhouette contours*, which are formed by the object and a background structure. Most organs are approximately convex, so self-occluding contours are rare events. By contrast, silhouette contours are common, and we propose to use these to constrain the organ's shape during registration. We require the silhouette contours to be provided in a keyframe image (Fig. 2). This is very difficult to automate because not all of the organ's boundaries in an image correspond to silhouette contours. Indeed, they are either silhouette contours, or contours formed by the silhouette of another structure occluding the organ. We illustrate this in Fig. 2 (top row) where abdominal fat occludes the posterior region of the uterus. The boundary between fat and the uterus conveys no information about the shape of the uterus. We therefore propose to extract silhouette contours with a fast semi-automatic process and a touch-screen interface. The keyframes are displayed and the operator traces the organ's silhouette contours with a finger stroke. We apply the method of [33] to snap the finger stroke to the nearest image contour, based on active contours attracted to the dominant image edge adjacent. This is very fast and a keyframe is processed in a matter of seconds.

D. Initialization

Initially, $\mathbf{x}_{\text{ref}}$ is initialized with a rigid transform $\mathtt{M}_a \in \mathcal{SE}_4$. If the laparoscope is in a canonical position w.r.t. the organ, $\mathtt{M}_a$ can be considered known *a priori*. Otherwise, we devised an interactive and relatively simple procedure to provide a rough estimate of $\mathtt{M}_a$. The method is based on solving PnP [34] between 3D points on the organ's surface model and their corresponding 2D points on one of the keyframes, selected interactively. The operator can first freely rotate the 3D model to present it from a similar viewpoint as the keyframe. Then they select at least 5 3D points on the 3D model and their corresponding points on the image. We found that a good strategy is to select the 4 equidistant points on the image along (but not on) the occluding contour and one roughly in the middle. The corresponding points on the 3D model can be guessed following the same pattern. Associating points from an untextured 3D model to its image is not in general a trivial task but we found that surgeons can easily perform the operation thanks to their detailed understanding of the anatomy. The correspondences are not required to be accurate, as this only serves as rough initialization of the initial registration.

To provide an initialization of the reconstruction scale factor s, we apply $\mathtt{M}_a$ to the model. We then use OpenGL to render a synthetic image of it using the calibrated intrinsic parameters of the laparoscope. OpenGL's z-buffer provides a depth map $d(x, y)$ from which we can compute s by comparing depths in d to depths in $\mathcal{Q}$. Specifically, for a 3D point $\mathbf{q}_j$, an estimate of s is $s_j = d(x_j, y_j)/\tilde{d}_j$, where $\tilde{d}_j$ is the depth of $\mathbf{q}_j$, and (x_j, y_j) its corresponding 2D point. The robust estimate is given by the median over all points, $s = \text{median}\{s_j\}$.

E. Energy-based Optimization

We describe the registration energy function and optimization process. To improve clarity we assume all image points in normalized camera coordinates, which is obtained from the intrinsic calibration, and thus define camera projection as $\pi([x, y, z]^\top) \stackrel{\text{def}}{=} [x, y]^\top / z$. Besides the model's internal energy $E_{\text{int}}(\mathbf{x})$, the energy function $E(\mathbf{x}, s) \in \mathbb{R}^+$ consists of two other terms. We introduce a point cloud data term E_{point} to help the organ's surface fit the reconstructed point cloud. To constrain the organ's silhouette contours to fit the silhouette contour fragments we also include a contour data term E_{con}. Thus, the energy $E(\mathbf{x}, s)$ is defined as:

$$E(\mathbf{x}, s) = E_{\text{point}}(\mathbf{x}, s; \mathcal{Q}) + \lambda_{\text{con}} E_{\text{con}}(\mathbf{x}, s) + \lambda_{\text{int}} E_{\text{int}}(\mathbf{x}), \quad (1)$$

where λ_{con} and λ_{int} are scalar weights (with defaults $\lambda_{\text{con}} = 100$ and $\lambda_{\text{int}} = 50$).

For the point cloud data term E_{point} we use an ICP-based energy term: it uses a set of *virtual point correspondences* $\mathcal{P} = \{\mathbf{p}_j\}$ with $|\mathcal{Q}| = |\mathcal{P}|$, where $\mathbf{p}_j \in \partial\Omega$ is the unknown position of point $\mathbf{q}_j$ on the organ's surface mesh Ω. For a given $(\mathbf{x}, s)$, E_{point} is computed by first transforming Ω according to $f(\cdot; \mathbf{x})$ and applying the estimated scale factor s to the point cloud, so that $\hat{\mathbf{q}}_j \leftarrow s\,\mathbf{q}_j$. Then, $\mathbf{p}_j$ is set to the closest point to $\hat{\mathbf{q}}_j$ on the surface's mesh. Similarly to the point-to-plane distance function of ICP with rigid objects, E_{point} uses a

robust point-to-plane distance function that allows the model to slide over the point cloud without resistance to improve convergence. The energy is as follows:

$$E_{\text{point}}(\mathbf{x}, s; \mathcal{Q}) = \frac{1}{M} \sum_{j=1}^{M} \rho \left(d_{\text{plane}} \left(v_j(\mathbf{x}), \hat{\mathbf{q}}_j \right) \right), \qquad (2)$$

where $v_j(\mathbf{x}) \in \mathbb{R}^4$ gives the organ surface's tangent plane at $f(\mathbf{p}_j)$. The function $d_{\text{plane}}(\mathbf{v}, \mathbf{q})$ gives the signed distance between a plane $\mathbf{v}$ and a 3D point $\mathbf{q}$. The function $\rho : \mathbb{R} \to \mathbb{R}^+$ is an *M-estimator*. The MVS reconstruction may contain points off the organ or poorly reconstructed; these are discarded by the M-estimator. We experimented with different types of estimators and found that a pseudo-L1 $\rho(x) \stackrel{\text{def}}{=} \sqrt{x^2 + \epsilon}$ offers good results (with default $\epsilon = 10^{-4}$).

Similarly, E_{con} uses virtual point correspondences on the organ surface mesh's occluding contours. More explicitly, for a given estimate $(\mathbf{x}, s)$ and a given keyframe i, we generate a set of virtual correspondences $\mathcal{R}_i = \{\mathbf{r}_1, \ldots, \mathbf{r}_{C(i)}\}$ containing, for each contour pixel $\mathbf{c}_k$, the unknown position $\mathbf{r}_k \in \partial\Omega$ of its corresponding 3D point on the model's surface. To compute the correspondence we first apply $f(\cdot; \mathbf{x})$ to the organ's surface mesh, and we bring the model in the reference frame of the camera i using $(\mathtt{R}_i, s\,\mathbf{t}_i)$. Then we render the surface mesh as described in §III-D and store all the pixels on the silhouette boundary in a set $\mathcal{B}$. Let $\mathcal{C}_i$ the set of all pixels belonging to the contour fragments in keyframe i. We compute for each contour pixel $\mathbf{c}_k \in \mathcal{C}_j$ its closest point $\mathbf{b}_k \in \mathcal{B}$. Finally, we set $\mathbf{r}_k$ as the 3D position of $\mathbf{b}_k$, computed using the render's depth buffer. From all the correspondences $\mathcal{R}_i$, E_{con} is computed as:

$$E_{\text{con}}(\mathbf{x}, s) = \frac{1}{C} \sum_{i=1}^{N} \sum_{\substack{\mathbf{c}_k \in \mathcal{C}_i \\ \mathbf{r}_k \in \mathcal{R}_i}} \rho \left(\left\| \pi \left(f(\mathbf{r}_k) \right) - \mathbf{c}_k \right\| \right), \qquad (3)$$

where C is the total number of contour fragment pixels. Similarly to E_{point}, we use the M-estimator ρ for robustness.

F. Optimisation

In order to improve convergence, we use a stiff-to-flexible strategy to optimize E. We start with a stiff model, we optimize E, and then we reduce the stiffness to account for more deformation. We use 6 stiffness levels, in which the value of λ_{int} is halved w.r.t. the previous level, *i.e.* $\lambda_{\text{int}}(l) = \lambda_{\text{int}}(l-1)/2$. At each level we alternate between computing the virtual correspondence sets ($\mathcal{R}_i$ and $\mathcal{P}$) and optimising E, via Gauss-Newton iterations with backtracking line search until either convergence or a maximum of 20 iterations is reached.

G. Real-time Tracking-by-Detection

1) Overview: Once the initial registration is computed, real-time tracking starts from the live feed of the laparoscope. Our approach updates the initial registration at each frame and is robust to common challenges including occlusions (*e.g.*, surgical tools), partial views and viewpoint changes.

2) Keypoint Mapping: For each keyframe, we render the 3D model and store the depth map of all the pixels lying within the model's silhouette. From this, the 3D position $\mathbf{u}$ of any 2D image keypoint located within the model's silhouette can be determined. For each keyframe, we then extract a set of image keypoint and to each one of them we associate its corresponding depth. We concatenate the keypoint from all images into a single list $\mathcal{F} = \{(\mathbf{u}_m, i_m, \mathbf{d}_m)\}$, where $\mathbf{u}_m \in \mathbb{R}^3$ is the 3D point associated to feature detected in the keyframe of index i_m, with $\mathbf{d}_m$ its associated descriptor.

The approach can be used with any keypoint detector, such as SIFT [3], ORB [35] or SURF [4]. We find that SIFT has better performances in image matching [36]. We used the recent open-source GPU implementation PopSift [37] to achieve real-time computation: for a typical HD 1920×1080 image of the uterus, up to 3000 keypoints can be found in about $11\,\text{ms}$ with a standard GPU. To significantly reduce the problem of wrongly tracking specularities, we detect saturated pixels with an intensity threshold of 250 and any keypoint within $5\,\text{pixels}$ to a saturated pixel is discarded.

3) Pose Estimation Overview: For each new image we extract a set of keypoints $\mathcal{G} = \{(\mathbf{y}_i, \tilde{\mathbf{d}}_i)\}$, where $\mathbf{y}_i$ is the 2D image point and $\tilde{\mathbf{d}}_i$ its descriptor. Our registration scheme has three main steps. First, a set of candidate matches between $\mathcal{F}$ and $\mathcal{G}$ is computed and then a pose hypothesis that best explains these matches is searched. Finally the best pose hypothesis is refined with the Levenberg-Marquardt (LM) algorithm by minimizing the reprojection errors.

4) Computing Candidate Matches: Good candidate matches of the sets $\mathcal{F}$ and $\mathcal{G}$ are those pairs with *(i)* strong descriptor agreement (*i.e.* low descriptor distance) and *(ii)* a low likelihood of being false. We achieve the latter condition by applying Lowe's Ratio Test (LRT) [3]. A novelty of using *multiple* keyframes is that we can also exploit match *coherence*. Specifically, consider a feature in the image that matches a feature in the i^{th} keyframe. It is likely to be correct if there exists other features that also match with features in the i^{th} keyframe. We adopt a "winner-takes-all" strategy to enforce coherence and reduce false matches. Let i^* be the index of the keyframe with the largest amount of candidate matches. Since SIFT is invariant to scale changes and image rotation, the keyframe i^* can be considered as the visually "closest" to the input image. we then recompute the candidate matches, but using *only* features from the keyframe i^*. Computational efficiency can be easily achieved as $\mathcal{F}$ is completely pre-computed and the distances to evaluate the descriptor likelihood are quite fast to compute on the GPU (*e.g.*, $\sim 3\,\text{ms}$ to match 2000 features from an HD image against 5600 features from 16 keyframes).

5) Computing 3D Pose: From the set of candidate matches, we find the best pose hypothesis that explains these matches using RANSAC from OpenCV's default implementation. In some images tracking may be impossible or unreliable, if the organ is not visible or very partially. A good indicator for deciding if pose can be estimated reliably is the number n_c of inlier matches found by RANSAC. If this is below a threshold (default 8 points), we reject the pose and consider the organ to be untrackable in that image. Finally, to reduce jitter, we feed

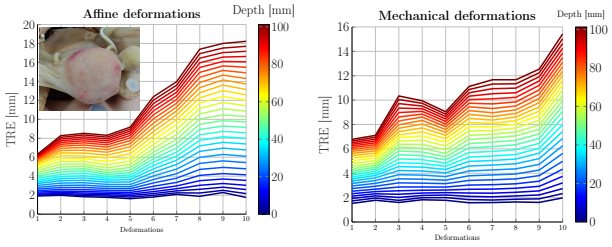

Fig. 4: TRE for affine (left) and mechanical (right) deformations. The deformation extent increases along the x-axis. The TRE is shown for different depths from the fundus.

the pose estimates into an Extended Kalman Filter (EKF).

6) Adding New Keyframes: The operator can manually add new keyframes to $\mathcal{F}$ on the fly during the live tracking. This mitigates any remaining jitter and stabilizes tracking if the organ is viewed from a viewpoint that is very dissimilar to the keyframes. Adding keyframes also copes with changes occuring between the reference and current intra-operative states in organ appearance (bleeding spots, change of color because of the clamping) and 3D orientation (due to mobilization). We add the keypoints of the current image and their associated 3D points to $\mathcal{F}$. We do not automate the process of adding new keyframes because without care it can lead to drift. Specifically, errors in the estimate pose then lead to incorrectly aligned keypoints in $\mathcal{F}$ and degradation in tracking accuracy. The operator can add the current frame as a new keyframe by pressing a bottom on the interface. A still image of the current frame with AR is then shown to the operator for validation.

IV. Experimental Validation

A. Overview

We first assess the Target Registration Error (TRE) on a phantom and we then present a thorough evaluation with offline videos of real surgeries[1], where we analyse three specific points of our pipeline: automatic keyframe selection, occlusion resistance in tracking and the use of new keyframes during tracking[2]. We then describe our experiences with deploying and running the pipeline in the real clinical setting. All the experiments are novel and more advanced compared to the workshop papers. We emphasize that all previous works on real-time markerless monocular laparoscopic registration with pre-operative organ models have only been evaluated with pre-recorded video data. Thus they have never made the step up to live tests in the OR and their robustness and practicality in the clinical setting have not been validated.

B. Quantitative Evaluation with Phantoms

In our workshop paper [27] we presented a thorough evaluation of the initial registration step with simulated data. A 3D uterus model was rendered with realistic camera motion and simulated initial deformations. We studied the influence of keyframe number on registration accuracy. The results showed that the registration error decreases with the number of keyframes, and saturates approximately at $1.4\,\mathrm{mm}$ for more than 8 keyframes. More generally, the error distribution tends to increase towards the cervix ($2\,\mathrm{mm}$ for 15 views up to $8\,\mathrm{mm}$ for 2 views) away from the uterus fundus (head), which is quite well constrained by the SfM point cloud as opposed to deeper regions near the cervix.

We present a new experiment to evaluate the TRE of the full registration pipeline (initial registration and tracking) with a realistic latex uterus phantom used for surgery training (Limbs and Things Inc. Model 60922, Fig. 4). We collected $N_{\mathrm{gt}} = 200$ images of the phantom, from which we computed a 3D surface model using MVS. We used this model as the preoperative model, and we synthetically deformed it to simulate different preoperative states with two kinds of deformations: isovolumetric affine deformations, and mechanical deformations [27] that simulate bending of the organ using a quadratic deformation law along a principal axis. We generated 10 examples of each type, progressively increasing in magnitude. The maximum amount of vertex displacement due to deformation is around 10 mm in both deformation types. For each deformation $\Theta(\cdot)$, we ran the full registration pipeline, using $N_{\mathrm{s}} = 15$ images for the initial registration step and $N_{\mathrm{gt}} - N_{\mathrm{s}}$ images for tracking. We discretized the model with a grid of $100 \times 100 \times 100$ voxels, with those interior to the uterus denoted by $\mathcal{V}$. We computed TRE for each voxel $\mathbf{q} \in \mathcal{V}$ in each image as $\|f(\Theta(\mathbf{q}), \mathbf{x}_t) - \mathtt{M}_{\mathrm{gt}}(\mathbf{q})\|$, where $\mathbf{x}_t$ are the model registration parameters estimated by our pipeline and $\mathtt{M}_{\mathrm{gt}}$ is the ground truth pose provided by the MVS reconstruction. TRE varies as a function of both the depth of the target and the amount of organ deformation. We visualize the trends in Fig. 4. We show TRE averaged over all the tracked frames for voxels at 30 different depths from the fundus surface, ranging from $0\,\mathrm{mm}$ to $100\,\mathrm{mm}$ (approximately the cervix). TRE at the fundus surface (depth $0\,\mathrm{mm}$) was below $2\,\mathrm{mm}$ for all deformations. In the near fundus (depth $< 20\,\mathrm{mm}$) TRE ranged from $1.8\,\mathrm{mm}$ to $3.9\,\mathrm{mm}$ for the most severe affine deformation. The error increased in depth towards the cervix, from $6.7\,\mathrm{mm}$ up to $15.6\,\mathrm{mm}$. The increase is normal and expected because of the large distance to the visible surface region. The low registration error in the fundus indicates the pipeline is sufficiently accurate to help locate uterine tumours in that region. Due to the large TRE at the cervix, AR for tumours in that region would be out of the intended use of the system.

C. Evaluation with Human Uteri in Pre-recorded Videos

In this section we present results on videos recorded during laparoscopic surgery. We test accuracy, computational complexity and the influence of the keypoint detector in our real-time organ tracking by detection system, named RT-OTD in the experiments. We demonstrate the robustness of our model-based approach w.r.t. classic SLAM approaches [35] and we show augmentation results, completing the AR pipeline

[1] All participants enrolled gave their written informed consent according to the approval of the ethical committee (IRB 2018-A03130-55).

[2] All supplementary videos are found at https://bit.ly/3lrUh5d indexed with capital letters from [A] to [I].

# frames		# poses	# matches	# matches winner	# inliers
5000	SURF	4569	406.16	46.94	32.93
	SIFT	4975	388.06	52.31	44.86
3029	SURF	2713	296.81	42.60	27.74
	SIFT	3029	294.00	64.87	52.81

TABLE I: Comparison in number of poses and matches between SIFT and SURF for two videos.

1) Automatic Keyframe Selection for 3D Reconstruction: We compare the automatic keyframe selection method *auto* to results obtained by sampling the exploratory video ($\sim 60\,\mathrm{s}$) to obtain the 15 keyframes, with different sampling methods: *equally* samples the video uniformly, *beginning*, *middle* and *end* take a keyframe every $t = 2\,\mathrm{s}$ from the beginning, around the middle and towards the end of the video, respectively. We then proceed with reconstruction and tracking for each strategy. Fig. 10 shows the 3D models. Qualitatively, *auto* gave a better model as all the others have many holes. This is explained by 1) some selected keyframes are blurry and 2) there are many similar keyframes, thus preventing a complete coverage of the full organ shape.

We recorded another video ($\sim 60\,\mathrm{s}$) in order to test the tracking using the 3D models obtained from each method. The video was purposely challenging, with the camera looking at parts of the uterus that were not completely reconstructed in any of the models, and no additional keyframes added during tracking. *auto* was able to track the largest number of frames (55.18 %), followed by *equally* (53.33 %), *middle* (52.48 %), *beginning* (46.14 %) and *end* (31.08 %). Fig. 11 reports tracking statistics. In general, *auto* recovers a larger number of valid matches, both w.r.t. all the keyframes in the database (Fig. 11.a) and the winning keyframe (Fig. 11.b) and computes the pose with a larger number of inliers. The reprojection error is also slightly better than all the other methods, especially considering the higher number of inliers over which it is computed. Overall, automatic keyframe selection improves the quality of reconstruction and tracking. Any sampling method is always potentially affected by motion blur; guiding the acquisition of the keyframe reduces the risk.

2) Comparison of Different Tracking Keypoints: We tested our tracking method RT-OTD using PopSift [37] and the SURF-GPU from OpenCV on two videos. We see from Table I that RT-OTD using SIFT establishes more camera poses (99.5 % and 100 % of the frames) compared to SURF (91.2 % and 89.6 %). Despite the fact that SURF recovers, on average, more matches between $\mathcal{F}$ and $\mathcal{G}$, SIFT has a higher discriminative power in selecting the winning keyframe i^*. As the table shows, the winning keyframe has, on average, more available matches from which RANSAC can sample to compute pose. This also results in more inliers found to support the computed pose. We show both pose components in Fig. 5 for the 5000 frame video. SIFT provides a more stable estimate for both as there are much fewer spikes in the estimates (spikes are typically incorrect estimates). This translates to a more stable motion estimate, improving overall AR quality with SIFT (see video [A]).

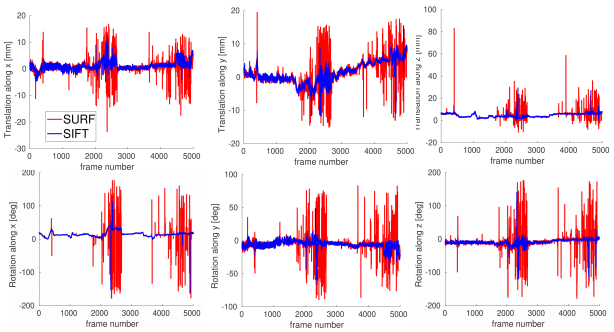

Fig. 5: Comparing SURF and SIFT on 5000 video frames.

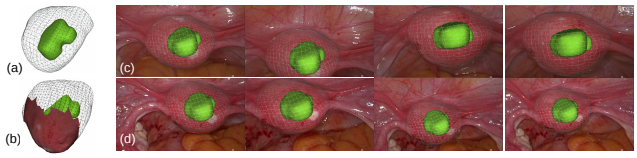

Fig. 6: (a) The 3D preoperative model of the uterus with the myoma as reconstructed from the MR, (b) the preoperative model registered with the intraoperative model obtained from the MVS reconstruction (c) some of the keyframes used for MVS with the AR augmentation (d) some frames of the video with the myoma shown as image overlay.

3) Tracking Accuracy and SLAM Evaluation: We evaluate our tracking stage using three human uteri captured before hysterectomy. This experiment is an update of the one presented in [28] with an evaluation of a state-of-the-art monocular SLAM approach (ORB-SLAM) [38]. SLAM approaches such as ORB-SLAM track a camera relative to a rigid 3D scene while simultaneously modelling the scene's 3D structure. ORB-SLAM is today the best SLAM system for use in laparoscopic surgery [35]. The uterus tends to dominate the field-of-view in uterine surgery, so a pertinent question is: would ORB-SLAM successfully track the camera with respect to the uterus? The answer is not obvious, because ORB-SLAM has built-in robustness that allows it to handle moving background structures.

The uteri are shown in Fig. 8 and we refer to them as U_1, U_2 and U_3. Each video includes around 500 frames showing image motion of the uterus due to motion induced by the cannula and camera motion. The uterus intra-operative surface is obtained with an exploratory video with 15 keyframes. The uterus body was marked by the surgeon with a bipolar grasper in $12 - 15$ different locations. This enabled us to generate Ground-Truth (GT) camera poses by tracking the obtained set of small marked regions ($\sim 3\,\mathrm{mm}$ in diameter). We show examples of these markers in Fig. 8. The marks were tracked using a small patch surrounding the image position of each marker, and fitted using a 2D affine transform. We verified all tracks and reinitialized them if they were lost. We then computed the marks' 3D positions and the uterus 3D poses in each frame using SfM. If fewer than four marks were visible in a frame we considered that the GT pose could not be estimated for that frame. We masked each mark so that the methods under comparison could not exploit the artificial texture each mark introduces. We computed the optimal scale factor for

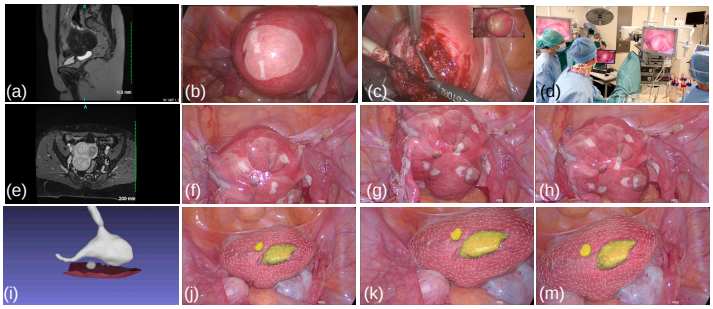

Fig. 7: Patient 1: (a) MR image showing one myoma, (b) myoma augmentation, (c) resection of the myoma, (d) deployment of our AR system in the OR. Patient 2: (e) MR image showing three myomas, (f,g,h) myoma augmentation. Patient 3: (i) two myomas segmented from the MR images, (j,k,m) augmentation of the myomas and the uterine cavity.

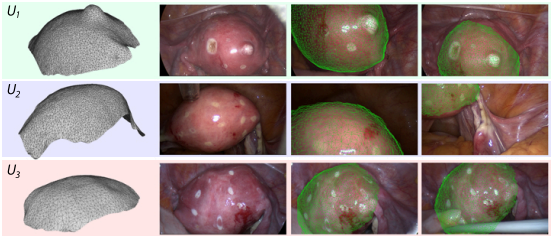

Fig. 8: The 3D models from MVS, the coagulation marks and the registered models.

each method w.r.t. GT. In Fig. 9 we summarize the results of

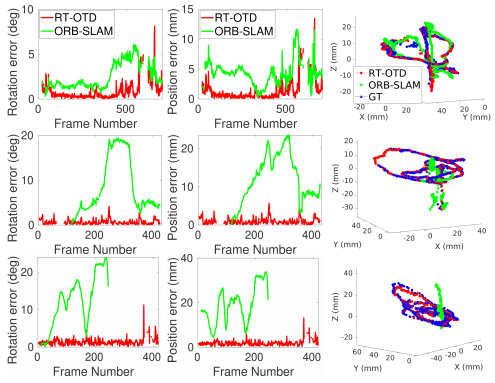

Fig. 9: Rotation error, position error and camera trajectory, for U_1, U_2 and U_3 from top to bottom.

RT-OTD using SIFT features. The rotation error (in degrees) and position error (in mm) are computed as the Euclidean norm between GT and estimates. There are gaps in the graphs when GT is not available. The results with RT-OTD are very accurate and stable in all cases, with an average position error of $2\,\mathrm{mm}$ and an average rotation error of $3°$. Fig. 9 also shows the tracking performed with ORB-SLAM [38]. While it tracks a large portion of the first sequence, it quickly degenerates and loses the track for the other two cases. This is because it builds a model with points from both the moving uterus and the background, violating the rigidity assumption. It is also affected by motion blur and when there are few matches.

4) Robustness Test for Tracking: Since it is hard to obtain GT data in the laparoscopic setting, we evaluated tracking robustness w.r.t. occlusions from the keyframes used for 3D

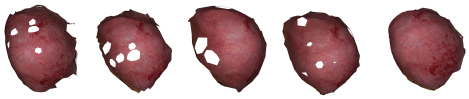

Fig. 10: Left to right: the 3D models generated using the keyframes from *beginning*, *middle*, *end*, *equally*, *auto*.

reconstruction, for which a reference pose is available from SfM, forming the GT. Using the data collected from 5 patients, we simulated occlusions and compared the pose computed by the tracker to the GT. In a first experiment, we used the known mask of the uterus to add a black occluder starting from the external contour and towards the center of the uterus. We gradually increased the size of the occluder and launched tracking on the keyframe. Fig. 11.e shows the overall percentage of frames that could be tracked against the occlusion ratio. The occlusion ratio is computed from the number of pixels of the occluder and of the uterus. The graph shows that tracking copes with up to $60\,\%$ occlusion, where $\sim 80\,\%$ of the frames are tracked. For the same experiment, Fig. 11.f shows that the pose rotational error is below $\sim 1°$ for up to $50\,\%$ occlusion, demonstrating good robustness. In a second experiment we simulated the presence of tools or bleeding covering the appearance of the uterus by growing black spots randomly distributed on the the uterus. Fig. 11.g shows that tracking successfully estimates pose for up to $60\,\%$ occlusion, where $\sim 90\,\%$ of the frames are tracked, with rotational error below or of the order of $\sim 1°$ (see videos [B,C]).

5) Additional Keyframes During Tracking: We present results to show the impact of adding keyframes during tracking. We refer the reader to the video material to observe the qualitative impact on the stability of pose and quality of AR (videos [D-F]). Table II shows the number of tracked keyframes for three videos with and without additional frames. Importantly, even a single new keyframe can greatly improve the number of tracked frames, as in video [D] in which it almost doubles the number of tracked frames. In general however, the number of tracked frames does not vary considerably but the tracking is much more stable.

6) Computational Times: Our hardware is composed of a desktop PC running Linux Ubuntu with an Intel Core i7-5960X CPU running at $3.00\,\mathrm{GHz}$ with $16\,\mathrm{GB}$ of RAM and

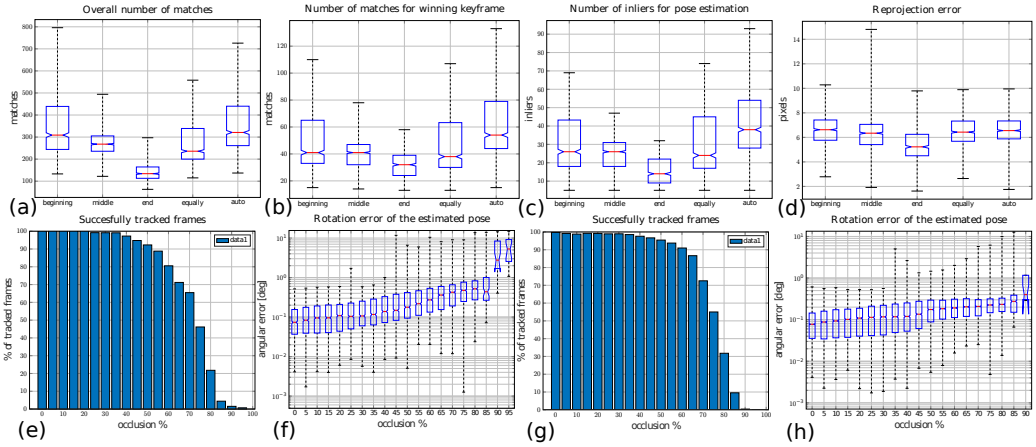

Fig. 11: (a,b,c,d) Tracking statistics (number of matches, inliers, and pose errors) for each keyframe sampling method. (e,f,g,h) From left to right, histogram of successfully tracked frames and rotation error of estimated pose, for both experiments.

video id	# frames	% tracked w/o additional keyframes	added keyframes	% tracked
20190801	1943	55.58 %	1	99.90 %
20190612	5119	95.64 %	8	97.15 %
20190724	2444	87.52 %	4	87.96 %

TABLE II: Number of tracked frames with and without adding new keyframes during tracking.

an NVIDIA GeForce GTX 980 Ti graphics card. On average, tracking takes 16.35 ms with a full HD 1920 × 1080 image. SIFT takes 11.1 ms, descriptor matching between $\mathcal{F}$ and $\mathcal{G}$ takes 3.21 ms and pose estimation takes 0.18 ms.

7) Augmentation Results: Fig. 6(a) shows the preoperative data of a patient whose uterus contains a myoma (in green) of a size of 11.3 mm × 22.9 mm × 17.5 mm. The preoperative mesh of the uterus has 2488 vertices and 4972 faces. From the exploratory video 15 keyframes were extracted to perform MVS reconstruction, obtaining a 3D model of 1760 vertices. Fig. 6(b) shows the successful alignment between the preoperative data and the MVS reconstruction. The cost function (1) was optimized in 6 iterations in approximately 21 s. In Fig. 6(c) we show four keyframes used for the reconstruction together with an overlay of the uterus surface registered to each frame. Qualitatively we see that the 3D preoperative model aligns well to the image of the uterus. Finally, Fig. 6(d) reports the visual augmentation of the myoma in some frames of a video recorded during surgery, clearly showing that the uterus is tracked well (see videos [G-I]).

D. Live Tests in the Operating Room

We tested our AR system in laparoscopic surgery [39] and report on three patients with one, three and two uterine myomas respectively. We built the 3D preoperative models of the organ and the myomas from preoperative T2-weighted MR. We used our system so that the surgeon could see the location of the myomas in real time. In Fig. 7(a) we show the MR of the first patient with a 6 cm uterine myoma. Fig. 7(b) shows the augmentation of the myoma and Fig. 7(c) shows its resection. At that stage our algorithm shows a past augmentation due to the changes in the uterus surface. In Fig. 7(d) we show our system deployed in the operating room. The MR of the second patient is shown in Fig. 7(e), showing three myomas that are visualized with AR in Fig. 7(f,g,h). The third patient had two myomas whose segmentation from MR is shown in Fig. 7(i). Examples of augmentation of the myomas using bright colors and the uterine cavity mesh are displayed in Fig. 7(j,k,m). We recall that this is the first time one demonstrates markerless registration and AR during *live* laparoscopic surgery. In our experience both the initial and tracking AR stages are valuable to the surgeon. The first stage is necessary to give the first appreciation of hidden structure locations (tumours and the uterine canal). Recall however that the surgeon is using a monocular camera. The value of the tracking stage is to give the sensation of depth and parallax, and to help orient the uterus to establish a good resection plane. Specifically, the surgeon can move the uterus with the cannula and they receive interactive visual feedback. Depth and spatial comprehension of hidden structures are easier because of motion and parallax effects.

V. Conclusion

The proposed framework is the first complete markerless real-time AR guidance system for laparoscopic surgery of the uterus. Its major advantage is that it does not require special hardware and works with a standard monocular laparoscope and an off-the-shelf computer. This enables a quick set-up in the OR, where no other hardware should interfere, perturb or distract the surgeon. Overall, the required steps for calibration, 3D reconstruction and registration take around ~ 10 min. Most of the interactive and manual operations, such as the masking and the contour annotations are carried out by an operator and can be done in parallel while the surgeon proceeds with other surgery tasks (*e.g.* clamping). Thus the impact on the clinical activity of the surgeon is limited to acquiring images for the calibration and the 3D reconstruction. Concerning AR and the visual feedback, the experiments show that the proposed approach for tracking is robust and responsive. It runs in real-time with very low jitter, further mitigated by

adding keyframes. One of the limitations of our system is that it relies on a few manual interactions. This will drive our research directions to further improve usability and accuracy. First is masking of the keyframes for 3D reconstruction. Even if the effort is mitigated by the touchscreen (overall time kept under a minute), automatically detecting the organ in the image would further the usability. We are investigating the use of CNNs to automatically extract the occluding contour fragments for initial registration [40], but while the first results are encouraging, some more efforts are required to robustly integrate it in the pipeline. As for the tracking part, an interesting and challenging research direction is to automate the addition of keyframes. This is non-trivial because of the potential for drift caused by the accumulation of small registration errors. Finally, we are preparing a follow-up clinical trial using a hybrid OR with interventional CT imaging to quantify TRE with patients in the OR, which is a notoriously difficult task, following the ideas proposed in [41]. When the tip of the laparoscope is in the CT field-of-view, laparoscopic and CT images can be registered with very high accuracy, enabling in-vivo end-to-end evaluation of target registration error.

REFERENCES

[1] A. Bartoli, T. Collins, N. Bourdel, and M. Canis, "Computer assisted minimally invasive surgery: Is medical computer vision the answer to improving laparosurgery?" *Medical Hypotheses*, vol. 79, no. 6, pp. 858–863, Dec. 2012.

[2] D. J. Mirota, M. Ishii, and G. D. Hager, "Vision-based navigation in image-guided interventions," *Ann. Rev. Biomed. Eng.*, vol. 13, no. 1, pp. 297–319, Aug. 2011.

[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[4] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *CVIU*, vol. 110, no. 3, pp. 346–359, 2008.

[5] S. Bernhardt, S. A. Nicolau, L. Soler, and C. Doignon, "The status of augmented reality in laparoscopic surgery as of 2016," *Medical Image Analysis*, vol. 37, pp. 66–90, apr 2017.

[6] G. Puerto-Souza, J. Cadeddu, and G. Mariottini, "Toward long-term and accurate augmented-reality for monocular endoscopic videos," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 10, pp. 2609–2620, 2014.

[7] H. Altamar, R. Ong, C. Glisson, D. Viprakasit, M. Miga, S. Herrell *et al.*, "Kidney deformation and intraprocedural registration: A study of elements of image-guided kidney surgery." *Endourology*, 2010.

[8] A. Amir-Khalili, M. S. Nosrati, J.-M. Peyrat, G. Hamarneh, and R. Abugharbieh, "Uncertainty-encoded augmented reality for robot-assisted partial nephrectomy." in *AECAI@MICCAI*, 2013, pp. 182–191.

[9] D. Cohen, E. Mayer, D. Chen, A. Anstee, J. Vale, G.-Z. Yang *et al.*, "Augmented reality image guidance in minimally invasive prostatectomy," in *Prostate Cancer Imaging. Computer-Aided Diagnosis, Prognosis, and Intervention*. Springer Berlin Heidelberg, 2010, pp. 101–110.

[10] G. Hamarneh, A. Amir-Khalili, M. Nosrati, I. Figueroa, J. Kawahara, O. Al-Alao *et al.*, "Towards multi-modal image-guided tumour identification in robot-assisted partial nephrectomy," in *MECBME*, 2014.

[11] N. Haouchine, J. Dequidt, I. Peterlík, E. Kerrien, M.-O. Berger, and S. Cotin, "Image-guided simulation of heterogeneous tissue deformation for augmented reality during hepatic surgery," in *ISMAR*, 2013.

[12] L.-M. Su, B. P. Vagvolgyi, R. Agarwal, C. E. Reiley, R. H. Taylor, and G. D. Hager, "Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3D-CT to stereoscopic video registration," *Urology*, vol. 73, no. 4, pp. 896–900, 2009.

[13] D. Stoyanov, M. V. Scarzanella, P. Pratt, and G.-Z. Yang, "Real-time stereo reconstruction in robotically assisted minimally invasive surgery," in *MICCAI*, 2010, pp. 275–282.

[14] X. Liu, A. Sinha, M. Unberath, M. Ishii, G. D. Hager, R. H. Taylor *et al.*, "Self-supervised Learning for Dense Depth Estimation in Monocular Endoscopy," in *Lecture Notes in Computer Science*, 2018, pp. 128–138.

[15] M. Visentini-Scarzanella, T. Sugiura, T. Kaneko, and S. Koto, "Deep monocular 3D reconstruction for assisted navigation in bronchoscopy," *Int. J. Comput. Ass. Rad.*, vol. 12, no. 7, pp. 1089–1099, jul 2017.

[16] N. Haouchine, J. Dequidt, M.-O. Berger, and S. Cotin, "Monocular 3d reconstruction and augmentation of elastic surfaces with self-occlusion handling," *IEEE Trans. Vis. Comput. Graphics*, vol. 21, no. 12, pp. 1363–1376, Dec. 2015.

[17] R. Plantefève, I. Peterlik, N. Haouchine, and S. Cotin, "Patient-specific biomechanical modeling for guidance during minimally-invasive hepatic surgery," *Ann. Biomed. Eng.*, vol. 44, no. 1, pp. 139–153, Aug. 2015.

[18] M. R. Robu, J. Ramalhinho, S. Thompson, K. Gurusamy, B. Davidson, D. Hawkes *et al.*, "Global rigid registration of CT to video in laparoscopic liver surgery," *Int. J. Comput. Ass. Rad.*, vol. 13, no. 6, pp. 947–956, May 2018.

[19] D. Reichard, D. Häntsch, S. Bodenstedt, S. Suwelack, M. Wagner, H. Kenngott *et al.*, "Projective biomechanical depth matching for soft tissue registration in laparoscopic surgery," *Int. J. Comput. Ass. Rad.*, vol. 12, no. 7, pp. 1101–1110, May 2017.

[20] M. S. Nosrati, J.-M. Peyrat, J. Abinahed, O. Al-Alao, A. Al-Ansari, R. Abugharbieh *et al.*, "Simultaneous multi-structure segmentation and 3D non-rigid pose estimation in image guided robotic surgery," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 1–12, 2016.

[21] J.-N. Brunet, A. Mendizabal, A. Petit, N. Golse, E. Vibert, and S. Cotin, "Physics-Based Deep Neural Network for Augmented Reality During Liver Surgery," in *Proc. of MICCAI*, 2019, pp. 137–145.

[22] M. Pfeiffer, C. Riediger, J. Weitz, and S. Speidel, "Learning soft tissue behavior of organs for surgical navigation with convolutional neural networks," *Int. J. Comput. Ass. Rad.*, vol. 14, no. 7, 2019.

[23] F. Mahmood, R. Chen, and N. J. Durr, "Unsupervised reverse domain adaptation for synthetic medical images via adversarial training," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2572–2581, Dec. 2018.

[24] S. Thrun, "Robotic Mapping: A Survey," in *Exploring Artificial Intelligence in the New Millenium*. Morgan Kaufmann, 2002, pp. 1–35.

[25] N. Mahmoud, I. Cirauqui, A. Hostettler, C. Doignon, L. Soler, J. Marescaux *et al.*, "ORBSLAM-Based Endoscope Tracking and 3D Reconstruction," in *Proc. CARE*, oct 2017, pp. 72–83.

[26] J. Lamarca, S. Parashar, A. Bartoli, and J. M. M. Montiel, "DefSLAM: Tracking and Mapping of Deforming Scenes from Monocular Sequences," aug 2019.

[27] T. Collins, D. Pizarro, A. Bartoli, M. Canis, and N. Bourdel, "Computer-assisted laparoscopic myomectomy by augmenting the uterus with pre-operative MRI data," in *ISMAR*, 2014.

[28] ——, "Realtime wide-baseline registration of the uterus in laparoscopic videos using multiple texture maps," *MIAR*, 2013.

[29] T. Collins, P. Chauvet, C. Debize, D. Pizarro, A. Bartoli, M. Canis *et al.*, "A system for augmented reality guided laparoscopic tumour resection with quantitative ex-vivo user evaluation," in *CARE*, 2017, pp. 114–126.

[30] AliceVision, "Meshroom: A 3D reconstruction software." 2018. [Online]. Available: https://github.com/alicevision/meshroom

[31] J.-Y. Bouguet, "Pyramidal implementation of the affine lucas kanade feature tracker," Intel Corporation, Tech. Rep. 4, 2001.

[32] AliceVision, "Photogrammetric Computer Vision Framework," 2017. [Online]. Available: https://alicevision.github.io/

[33] E. N. Mortensen and W. A. Barrett, "Intelligent scissors for image composition," in *SIGGRAPH*, 1995, pp. 191–198.

[34] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An Accurate O(n) Solution to the PnP Problem," *Int. J. Comput. Vision*, vol. 81, no. 2, pp. 155–166, feb 2009.

[35] N. Mahmoud, T. Collins, A. Hostettler, L. Soler, C. Doignon, and J. M. M. Montiel, "Live tracking and dense reconstruction for handheld monocular endoscopy," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 79–89, Jan. 2019.

[36] T. Tuytelaars and K. Mikolajczyk, "Local Invariant Feature Detectors: A Survey," *Found. Trends Comput. Graph. Vis.*, vol. 3, no. 3, 2007.

[37] C. Griwodz, L. Calvet, and P. Halvorsen, "PopSift: A faithful SIFT implementation for real-time applications," in *Proc. ACM Multimedia Systems Conference*. ACM, 2018, pp. 415–420.

[38] R. Mur-Artal and J. D. Tardos, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, p. 12551262, Oct 2017.

[39] N. Bourdel, T. Collins, D. Pizarro, C. Debize, A.-s. Grémeau, A. Bartoli *et al.*, "Use of augmented reality in laparoscopic gynecology to visualize myomas," *Fertility and sterility*, vol. 107, no. 3, 2017.

[40] T. François, L. Calvet, S. M. Zadeh, D. Saboul, S. Gasparini, P. Samarakoon *et al.*, "Detecting the occluding contours of the uterus to automatise augmented laparoscopy: score, loss, dataset, evaluation and user study," *Int. J. Comput. Ass. Rad.*, May 2020.

[41] S. Bernhardt, S. A. Nicolau, A. Bartoli, V. Agnus, L. Soler, and C. Doignon, "Using shading to register an intraoperative CT scan to a laparoscopic image," in *CARE*. Springer, Cham, 2016, pp. 59–68.