Combining Conformal Deformation and Cook-Torrance Shading for 3D Reconstruction in Laparoscopy

Abed Malti and Adrien Bartoli
ALCoV-ISIT
UMR 6284 CNRS / Université d'Auvergne
Clermont-Ferrand, France

***Abstract*—We propose a new monocular 3D reconstruction method adapted to reconstructing organs in the abdominal cavity. It combines both motion and shading cues. The former uses a conformal deformation prior and the latter the Cook-Torrance reflectance model. Our method runs in two phases: first, a 3D geometric and photometric *template* of the organ at rest is reconstructed in-vivo. The geometric shape is reconstructed using RSfM (Rigid Shape-from-Motion) while the surgeon is exploring – but not deforming – structures in the abdominal cavity. This geometric template is then used to retrieve the photometric properties. A non-parametric model of the light's direction of the laparoscope and the Cook-Torrance reflectance model of the organ's tissue are estimated. Second, the surgeon manipulates and deforms the environment. Here, the 3D template is conformally deformed to globally match a set of few correspondences between the 2D image data provided by the monocular laparoscope and the 3D template. Then the coarse 3D shape is refined using shading cues to obtain a final 3D deformed shape. This second phase only relies on a single image. Therefore it copes with both sequential processing and self-recovery from tracking failure.**

The proposed approach has been validated using *(i)* ex-vivo and in-vivo data with ground-truth, and *(ii)* in-vivo laparoscopic videos of a patient's uterus. Our experimental results illustrate the ability of our method to reconstruct natural 3D deformations typical in real surgical procedures.

***Index Terms*—Laparoscopy, monocular 3D reconstruction, deformable surface, shading, motion.**

I. Introduction

The problem of 3D reconstruction in monocular laparoscopy has recently become a field of promising research. This has been made possible thanks to recent advances in 3D reconstruction of deformable surfaces [1], [2], [3], [4] and the extraordinary potential that such techniques can offer to new view point synthesis, augmented reality with 3D pre-operative data (MRI, CT, etc) and surgery planning, to name a few. However, if these techniques have shown effectiveness in a well-controlled context, the peritoneal tissues present three main difficulties: *(i)* non-rigid motion, *(ii)* non-Lambertian reflectance, and *(iii)* lack of texture. To be able to use motion and photometry for peritoneal tissues, a model of its parameters for the non-rigid motion and the BRDF (Bidirectional Reflectance Distribution Function) have to be estimated. It is clear that the estimation of the mechanical and photometric properties has to be done in-vivo since these parameters change across patients. These properties make 3D shape recovery from monocular laparoscopy a difficult and open problem. On the one hand, DSfM (Deformable-Structure-from-Motion) has shown effectiveness in recovering 3D shape after elastic deformations in laparoscopy [5], [4]. However, with these methods the 3D shape may be quite sparse. Human organs are usually textureless and very specular. This makes it difficult to densely cover their deforming surface with feature correspondences using automatic feature detection and matching. On the other hand, SfS (Shape-from-Shading) allows one to recover surface details. However, it is difficult in practice because the reflectance of the organ tissues is complex and the SfS problem has been mostly solved for Lambertian surfaces [6]. In addition, SfS does not allow one to solve temporal registration between successive images. In order to take advantage of DSfM and SfS and overcome their drawbacks, we propose to combine them in a DSfMS (Deformable-Shape-from-Motion-and-Shading) framework. This paper is an improvement of our former work [7] where we proposed a combination of motion and shading cues but we assumed a Lambertian reflectance model in the shading part. In this new work, we propose to use the Cook-Torrance reflectance model [8]. We prove with both qualitative and quantitative results that this assumption fits better the reflectance model

of the tissues than the Lambertian or the Oren-Nayar [9] models.

Paper organization. Section II presents state-of-the-art. Section III gives an overview of our DSfMS. Section IV presents the reconstruction of the photometric template using the Cook-Torrance reflectance model. Section V presents our 3D reconstruction method based on motion and shading cues. Section VI reports experimental results. Our notation will be introduced throughout the paper.

II. Related Work and Contribution

The various methods of 3D sensing in laparoscopy can be classified as active and passive [10]. As active approaches, [11], [12] have proposed a technique based on the detection of a laser beam line. In [13] a prototype of Time-of-Flight (ToF) endoscope was designed. If these approaches offer 2.5D views (depth maps) of the current image they do not solve the registration problem. Solving this problem is required for important applications such as augmented reality. In passive approaches both stereo and monocular endoscopes are concerned. In [5], [14] methods based on disparity map computation for stereo-laparoscope have been proposed. Visual SLAM for dense surface reconstruction has been proposed in [15]. In monocular DSfM approaches, the computer vision community has made important achievements in template-based 3D reconstruction. Template-based methods provide a dense surface recovery rather than just a sparse one as in the previously cited methods. This allows one to render the surface from a new viewpoint and opens applications based on augmented reality. Template-based monocular 3D recovery needs priors to have a unique consistent solution. Different types of physical and statistical priors were proposed [1], [2], [3]. Recently a 3D conformal method has been proposed to reconstruct elastic deformations in the context of laparosurgery [4]. To provide good reconstruction results, DSfM needs feature detection and matching in the deformed areas which are not easy to obtain because of the textureless nature of some tissues. Alternatively, SfS is a 3D reconstruction method which does not need feature correspondences [16], [6]. Recovering depths using shading cues has been extensively used for both rigid and deformable objects [16]. This method i quite effective with perfect diffuse surfaces (Lambertian reflectance model), but fails for surfaces which present specular reflections.

To overcome the bottleneck of SfS and DSfM, we can take advantage of both methods: we use feature-based 3D reconstruction to recover a coarse deformed 3D surface and we use shading to refine the reconstruction of areas which lack feature correspondences. This combined approach has been used in several other conditions to recover coarse to fine 3D shapes. For instance in rigid 3D reconstruction [17] presented an algorithm for computing optical flow, shape, motion, lighting, and albedo from an image sequence of a rigidly-moving Lambertian object under distant illumination. [18] proposed an approach to recover shape details in a dynamic scene captured with a multi-camera setup.

This paper is based on our recent work [7] where we proposed a method to combine motion and shading cues assuming a Lambertian model. In our present work, we improve this approach by using the Cook-Torrance reflectance model. Indeed, this function is known to better represent the physical model of the diffuse/specular reflectance properties of surfaces. We experimentally show with both qualitative and quantitative results that the Cook-Torrance model combined with motion cues performs better than the Lambertian model combined with motion cues as used in [7]. We take advantage of our recent work on estimating the Cook-Torrance parameters [19] to have fully automated 3D reconstruction.

Contribution. The contributions of our work are three folds. *(i)* Enhancing the 3D geometric template [4] with a photometric template by estimating the Cook-Torrance reflectance parameters from in-vivo images. *(ii)* Combining motion and shading cues, with a realistic reflectance model to recover the 3D deformed tissue. The motion cues take advantage of a few point correspondences to recover a coarse deformation of the tissue and the shading cues take advantage of the estimated Cook-Torrance parameters to accurately refine the 3D reconstruction. *(iii)* Our proposed 3D reconstruction method is qualitatively and quantitatively compared to two previous methods: the first is based only on motion cues with conformal priors [4] and the second is based on combining motion and shading cues but assumes a Lambertian model of reflectance [7].

III. Overview of DSfMS

As depicted in Figure 1, our DSfMS system has two main phases:

1) Template reconstruction. In this phase both the 3D structure and the Cook-Torrance parameters are recovered, by assuming that the scene remains approximately rigid as the surgeon explores it with

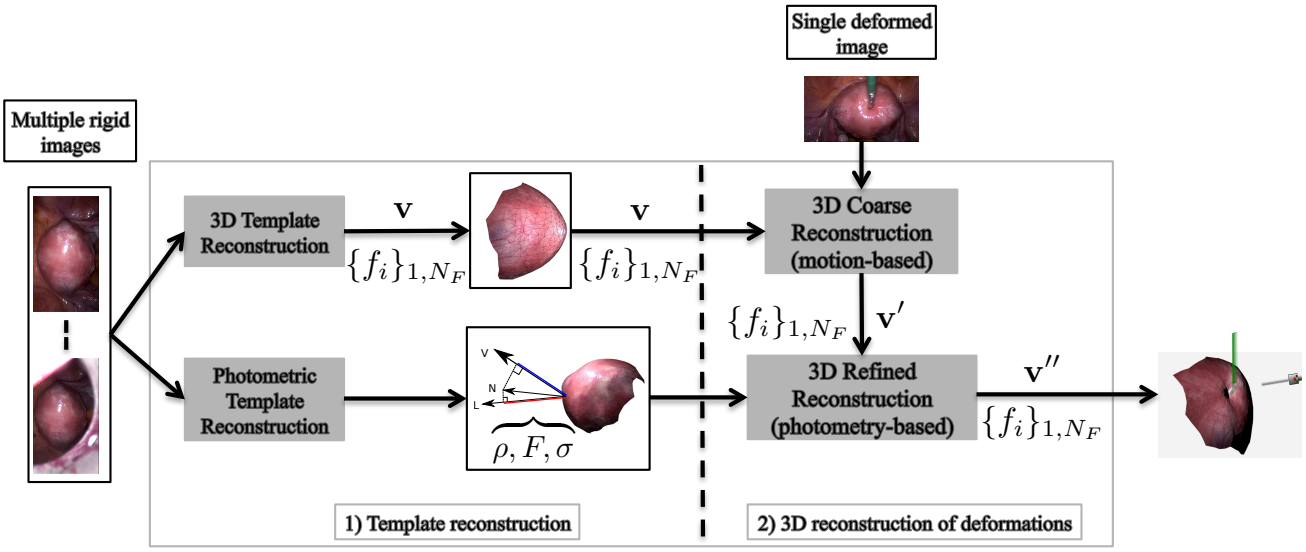

Fig. 1. Principle of our DSfMS (Deformable Shape-from-Motion-and-Shading) approach. In the first phase the surgeon explores the abdominal cavity without deforming it; RSfM (Rigid Shape-from-Motion) is used to find the 3D shape called the 3D template (N vertices and N_F faces). This 3D shape is used to infer the Cook-Torrance reflectance parameters and calibrate the light directions. In the second phase, the 3D template is used to infer the 3D shape deformed as observed from only a single laparoscopic view. This makes the approach resistant to registration and tracking errors and well-adapted to live sequential processing.

the laparoscope. Using a calibrated camera and the 5-point algorithm for RSfM [20], a 3D point cloud representing the organ's shape is reconstructed. The 3D point cloud is then meshed to provide a dense 3D surface, parameterized on the 2D plane via conformal flattening [21]. The procedure of reconstruction of the 3D geometric template is explained with more details in [4]. This geometric shape is then used to estimate the Cook-Torrance reflectance model (c.f. section IV).

2) 3D reconstruction of deformations. The surgeon is now free to proceed and manipulate the target surface, and consequently induces non-rigid deformations with the surgery tools. Here, the template reconstructed in phase 1) is used to perform 3D reconstruction from raw laparoscopic images. The 3D shape is computed by globally deforming the template assuming conformal deformations and then refined using shading cues with the estimated Cook-Torrance parameters (c.f. section V).

IV. Photometric Template Reconstruction

The Cook-Torrance model is known as being one of the most meaningful physical representation for complex surface reflectance modelling [8]. If we assume a linear radiometric response of the camera sensor, then the predicted image intensity $\hat{I}$ with this model depends on three vectors: the shape normal $\mathbf{N}$, the viewing direction $\mathbf{V}$ and the light direction

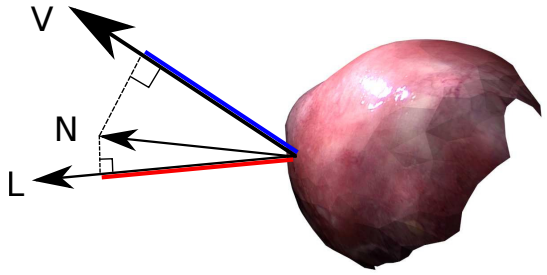

Fig. 2. Geometry of light reflection with the Cook-Torrance model. The image irradiance depends on *(i)* the projection of the surface normal $\mathbf{N}$ onto the viewing direction $\mathbf{V}$ and *(ii)* on the projection of $\mathbf{N}$ onto the incident light direction $\mathbf{L}$. The surface roughness D, the Fresnel parameter F and the albedo ρ are assumed constant.

$\mathbf{L}$ (see Figure 2 for a local geometry representation of the reflection). It is given by:

$$\hat{I}(\rho, F, \sigma, \mathbf{N}, \mathbf{L}) = \underbrace{\frac{\rho}{\pi}(\mathbf{N}\cdot\mathbf{L})}_{\text{diffuse reflectance}} + \underbrace{\frac{F\ D(\sigma,\mathbf{N},\mathbf{L})}{\pi\ (\mathbf{N}\cdot\mathbf{L})\ (\mathbf{N}\cdot\mathbf{V})}}_{\text{specular reflectance}} \quad (1)$$

where

$$D(\sigma,\mathbf{N},\mathbf{L}) = \frac{1}{\sigma^2\cos^2\alpha}\exp\left(-\frac{\tan^2\alpha}{\sigma^2}\right) \quad (2)$$

The diffuse reflectance is assumed to be Lambertian and ρ is the diffuse albedo. The Fresnel coefficient F represents the refractive index of the tissue. The facet slope distribution D can be represented by the Beckmann distribution explicited in equation (2), where σ represents the roughness of the tissue

(small for mirror-like surfaces). $\alpha = \widehat{(\mathbf{N}, \mathbf{H})}$ is the angle between the normal and the bisector $\mathbf{H}$ of the angle $\widehat{(\mathbf{L}, \mathbf{V})}$. For estimating the Cook-Torrance reflectance parameters (ρ, F, σ), of a given tissue, reflections are first measured under various viewing and illumination angles. At this step, the 3D geometric template of the tissue is assumed to be computed as described in section III. The measured image irradiance I (the intensity of the image) can be then approximated by the prediction formula (1). If we assume that, near the specularities, the light direction is such that the bisector $\mathbf{H}$ is parallel to the surface normal, we can estimate the Cook-Torrance parameters by minimizing the difference between the measured and the predicted intensity [19]:

$$(\rho, F, \sigma) = \underset{\rho,F,\sigma}{\operatorname{argmin}} \sum_{i \in \tilde{\mathcal{S}}} \left(I(u_i, v_i) - \hat{I}(\rho, F, \sigma, \mathbf{N}_i, \mathbf{L}_i) \right)^2 \quad (3)$$

where $\tilde{\mathcal{S}}$ is the set of pixels that are neighbours of the specular pixels (taken over all the considered rigid images of the organ) and which are not saturated pixels. $I(u_i, v_i)$ is the measured intensity, $\mathbf{N}_i$ is the normal of surface organ and $\mathbf{L}_i$ is the light direction at these pixels. Notice that ρ and F are estimated up to scale representing the camera response factor and the light intensity. Once these parameters are evaluated, we use them to estimate the light directions over all the pixels of the laparoscopic image. This estimation assumes that the source light is rigidly attached to the camera body. The computation criterion is as follows:

$$\mathbf{L}_i = \underset{\mathbf{L}}{\operatorname{argmin}} \sum_{j \in \mathcal{M}} \left(I^j(u_i, v_i) - \hat{I}(\rho, F, \sigma, \mathbf{N}_i^j, \mathbf{L}_i) \right)^2 \quad (4)$$

where j is the image index within the set $\mathcal{M}$ of rigid images and i runs over the image pixels to assign a light direction to each.The global minimum of criterion (3) is computed using Branch-and-Bound and Second Order Cone Programming (SOCP) [19]. The minimization of criterion (4) is done using the Levenberg-Marquardt algorithm [19].

V. Monocular Conformal DSfMS

A. Coarse 3D Reconstruction

We use a triangular mesh representation of the surface. When conformally deformed, each triangle may undergo stretching or shrinking by penalizing changes in angles. In [4] a discrete quasi-conformal reconstruction of deformable surfaces is proposed from N_c point correspondences between the imaged deformed shape and the 3D template. In this work we propose to formalize the conformal penalty by directly minimizing the changes in angles. This formulation has the advantage of being independent of any extra hyper-parameter as the shearing and the scaling. In the template, the correspondences are given by their barycentric coordinates $(f_i \ \mathbf{b}_i)^\top$, $i = 1, \ldots, N_c$. f_i is the index of the triangle and $\mathbf{b}_i = (b_i^1, b_i^2, b_i^3)$ are the values of the barycentric coordinates. (u_i, v_i), $i = 1, \ldots, N_c$ are the corresponding pixels in the image where the deformed shape is projected. Extensible 3D reconstruction is formulated as:

$$\begin{aligned} \mathbf{v}' = \underset{\mathbf{v}'}{\operatorname{argmin}} & \underbrace{\sum_{i=1}^{N_c} \| \ \Pi \ (\mathbf{v}'(f_i)\mathbf{b}_i^\top) - (u_i, v_i) \ \|}_{\text{(reprojection cond.)}} \\ & + \lambda_1 \underbrace{\sum_{i=1}^{N} \sum_{j \in \mathcal{N}(\mathbf{v}_i)} (\alpha_i(\mathbf{v}) - \alpha_i(\mathbf{v}'))^2 + (\beta_i(\mathbf{v}) - \beta_i(\mathbf{v}'))^2}_{\text{(conformal energy)}} \\ & + \lambda_2 \underbrace{\| \ \Delta \mathbf{v}' \ \|^2}_{\text{(smoothing)}} \end{aligned} \quad (5)$$

where Π is the projective mapping from 3D to 2D including the intrinsics of the camera. $\mathbf{v}'$ is the matrix of vertices that represents the coarse reconstructed shape with motion cues (see Figure 1). $\mathbf{v}'(f_i)$ is the (3×3) matrix whose columns are the 3D coordinates of the vertices of face i. $\alpha_i(\mathbf{v})$ and $\beta_i(\mathbf{v})$ are two template angles of triangle f_i. $\alpha_i(\mathbf{v}')$ and $\beta_i(\mathbf{v}')$ are their corresponding angles in the deformed shape. The conformal energy term allows triangles to stretch but penalizes changes in angles. The smoothing energy term is expressed through the linear Laplace-Beltrami discrete linear operator Δ of dimension $N \times N$ [22], where N is the number of vertices in the 3D template mesh. λ_1 and λ_2 are real positive weights that tune the amount of penalty for the conformal and the smoothing energy terms. Their values are respectively set to 0.11 and 0.30 using the method described by [23]. This is further discussed in section VI-D.

B. Fine 3D Reconstruction

The resulting deformed shape with the set of vertices $\mathbf{v}'_i$, $i = 1, \ldots, N$, recovered from the previously described method can be refined using shading cues to obtain the final mesh $\mathbf{v}''_i$, $i = 1, \ldots, N$ (last box in the pipeline of Figure 1). Using the reconstructed photometric template, we formulate

3D reconstruction with motion and shading cues as:

$$\mathbf{v}'' = \underset{\mathbf{v}''}{\operatorname{argmin}} \underbrace{\sum_{i=1}^{N_c} \parallel (0\ 0\ 1)^\top \left(\mathbf{v}''(f_i) - \mathbf{v}'(f_i)\right) \mathbf{b}_i^\top \parallel}_{\text{(boundary cond.)}} + \lambda_6 \underbrace{\parallel \Delta \mathbf{v}'' \parallel^2}_{\text{(smoothing)}}$$
$$+ \lambda_3 \underbrace{\sum_{i=1}^{N} \parallel \Pi(\mathbf{v}_i'') - \Pi(\mathbf{v}_i') \parallel^2}_{\text{(reprojection cond.)}} + \lambda_5 \underbrace{\sum_{i \in \mathcal{S}_v} \parallel \mathbf{H}_i \times \mathbf{N}_i'' \parallel^2}_{\text{(specular vertices)}}$$
$$+ \lambda_4 \underbrace{\sum_{i \in \mathcal{D}_v} \parallel I(u_i, v_i) - \left(\frac{\mathbf{N}_i'' \cdot \mathbf{L}_i}{\pi} \rho + \frac{F\, D(\sigma, \mathbf{N}_i, \mathbf{L}_i)}{\pi (\mathbf{N}_i'' \cdot \mathbf{L}_i)(\mathbf{N}_i'' \cdot \mathbf{V})} \right) \parallel^2}_{\text{(diffuse vertices)} \quad (6)}$$

where $\mathcal{D}_v$ and $\mathcal{S}_v$ are respectively the diffuse and specular pixels which belong to the organ's tissue. A tool/tissue segmentation using graph cuts [24] allows us to determine these pixels. The specular pixels can be easily detected as saturated regions in the deformed image intensity I. $\mathbf{H}_i$ is the bisector written as $\mathbf{H}_i = \frac{\mathbf{V}+\mathbf{L}_i}{2}$. The real parameters λ_3, $\lambda_4, \lambda_5, \lambda_6$ are experimentally set to 0.21, 0.21, 0.21 and 0.17. Paragraph VI-D in the experimental results section discusses this choice. Through the boundary condition, this formulation gives confidence to the depth of the correspondences reconstructed by the conformal method using motion. The reprojection condition constrains the refinement of the vertices along the camera sightlines. The diffuse condition refines the diffuse vertices according to the Cook-Torrance model using shading. The specular vertices are constrained to have their normals parallel to the bisector direction of the source light and the viewing direction. Due to noise in the image intensity a smoothing term is needed to avoid bumpy surfaces. The diffuse and specular terms allow us to recover the deformed surface in regions where the data correspondences are missing.

VI. Experimental Results

Using ex-vivo and in-vivo animal and human data with ground-truth, our proposed method **MoT-CT** (**MoT**ion-**C**ook-**T**orrance) is quantitatively compared to three methods: **MoT** (**MoT**ion) which is based only on motion cues with conformal priors [4], **MoT-LaM** (**MoT**ion-**LaM**bertian) which is based on combining motion and shading cues but assumes a Lambertian model of reflectance [7] and **MoT-ON** (**MoT**ion-**O**ren-**N**ayar) which uses the Oren-Nayar [9] model instead of the Cook-Torrance Model. We also present some qualitative reconstructions using in-vivo images of a human uterus. The considered deformations range from 2 mm to 12 mm per mesh vertex up to a rigid transform of the mesh. This interval of deformation is acceptable for early surgery step when the surgeon deforms the organ to find the locations of abnormal tissues.

The proposed method is implemented and tested with Matlab_R2013a running on a MAC OS X 10.8 system with an Intel Core 2 Duo CPU running at 2.26 GHz. The template reconstruction takes about 15 seconds. The light and Cook-Torrance calibration take about 10 seconds. These two steps are processed once at the beginning of the experiments or surgery. The 3D reconstruction of deformations lasts about 10 seconds.

A. Ex-Vivo Data With Ground-Truth

In order to acquire real ex-vivo datasets we used two laparoscopes fixed through two trocars mounted on a pelvitrainer (see Figure 3). The two laparoscopes are mounted to two PointGrey Flea2 color cameras with two c-mounts. The two cameras are synchronized at 15 fps with a resolution of 1024×728 pixels. The distance between the two cameras is about 3 cm and the angle between the camera axes is about 30 degrees. This setup allows us to build reference ground-truth 3D models of ex-vivo organs with stereo views. The stereo reconstruction was about 0.2 mm. In order to avoid interference with the room light, we covered the pelvitrainer with black clothes and the room light was turned off during the experiments

In our validation with ex-vivo organs we use the lung and the liver of a lamb. In the first exploratory step we reconstruct the 3D template of these organ's tissues as shown in Figure 4. The Cook-Torrance parameters and the light are then calibrated as described in IV. In the deformation step, the lung and the liver are deformed with a surgery tool. A set of 600 deformed image frames are taken. We use on average a set of 15 point correspondences between template and deformed images. For the liver, we drew a set of patterns with surgery pen because of the extreme textureless aspect of its tissue. The correspondences were generated using SIFT [25]. Outliers and points outside the organs were removed by the method proposed by [26]. In Figure 5 we show a subset of different 3D reconstructions using our method from single views for different amounts of extensibility and curvature change with respect to the template. We can see that globally our method gives meaningful 3D reconstructions according to the deformed images.

B. In-Vivo Data With Ground-Truth

To obtain in-vivo datasets with ground-truth we

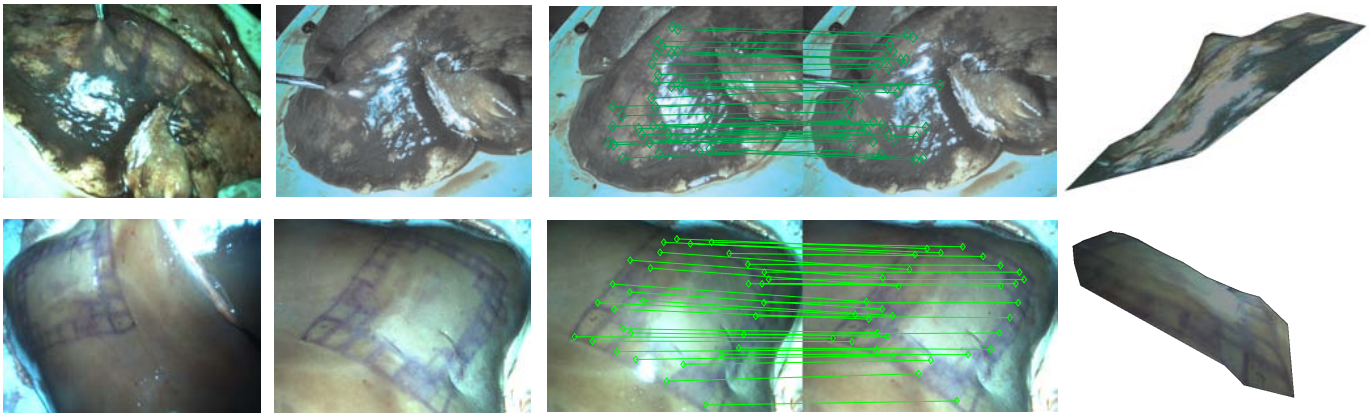

Fig. 5. Ex-vivo datasets: 3D reconstruction from a monocular laparoscope using our DSfMS method. First column: Left image from stereo view used to compute ground-truth deformation. Second column: Right image from stereo view. This image is used together with the left image to generate ground-truth 3D reconstruction. It is also used as single image to obtain 3D reconstruction with our method. Third column: correspondences between the template image and right image used for the 3D reconstruction with our method. Fourth column: 3D reconstruction with our method from single image. Quantitative 3D errors of reconstruction are shown in Table I.

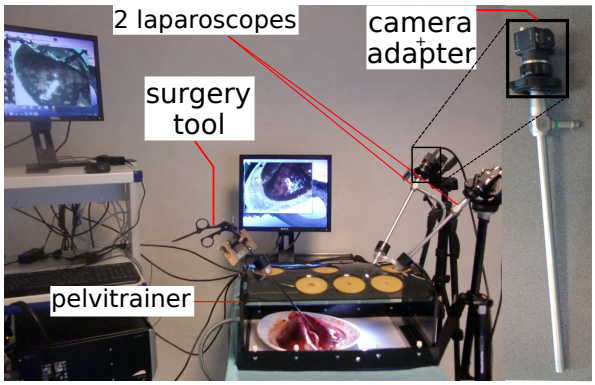

Fig. 3. Experimental setup to acquire real ex-vivo datasets. Two Pointgrey cameras are synchronized to obtain reference ground-truth data using stereo-views.

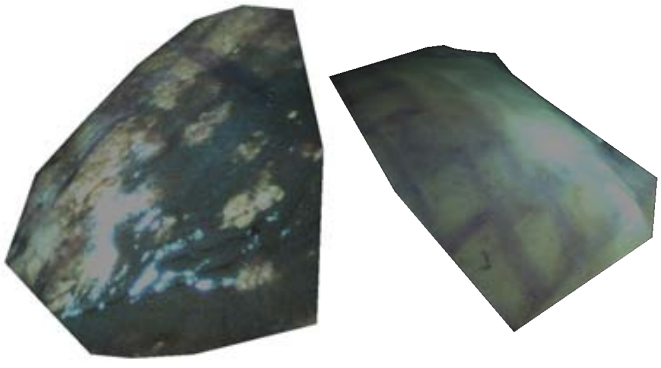

Fig. 4. Ex-vivo datasets: 3D template of the lung and the liver. The size of the box bounding the 3D shape of the lung is $45.40 \times 47.50 \times 20.16$ mm^3. The size of the box bounding the 3D shape of the lung is $42.40 \times 43.50 \times 18.16$ mm^3.

use two synchronized laparoscopes in a stereo setup to explore and deform the abdominal cavity of a living pig. The experiment is done in the Centre International de Chirurgie Endoscopique[1] under respect of ethical constraints. The laparoscopes and the synchronization framework follow the same setup as described before to acquire real ex-vivo datasets. However, to cope with the difficulty of having a non-constant rigid transform (stereo-transform) between the two laparoscopes we put a reference checkerboard inside the abdominal cavity. This checkerboard allows us at any frame to calibrate the stereo-transform from the left and right views to obtain ground-truth 3D information. The distance between the two cameras is about 3 cm and the angles between the camera axis is about 30 degrees. The stereo reconstruction was about 0.3 mm. In the first exploratory step we reconstruct the 3D template of three different organ's tissues: the bladder, the pericardium and the left lung. The obtained shapes are shown in Figure 6. The Cook-Torrance parameters and the light are then calibrated as described in IV. In the deformation step, the bladder and the pericardium are deformed with the checkerboard tool. The left lung also deforms due to the breathing. A set of 500 deformed image frames are taken for each tissue. For our reconstruction method we use on average a set of 20, 25 and 10 point correspondences respectively for the bladder, the pericardium and the left lung. They were generated using SIFT [25]. Outliers and points outside the organs in concern were removed by the method proposed in [26]. In Figure 7 we show a subset of

[1]http://www.cice.fr/

different 3D reconstructions using our method from single views for different amounts of extensibility and curvature change with respect to the templates. We can see that globally our method gives meaningful 3D reconstructions according to the deformed images. The quantitative results are shown in table I and discussed in paragraph VI-E.

C. Surgery In-Vivo Data with Ground-truth

To validate the proposed approach on real in-vivo data, we use in-vivo sequences of a human uterus acquired using a monocular Karl Storz laparoscope. The frames are acquired at 25 fps and have a resolution of 1920×1080. The 3D template of the uterus is generated during the laparosurgery exploration step as previously described. The Cook-Torrance parameters and the light are then calibrated as described in IV. Deformations on the uterus are performed by a surgery tool. To construct the ground-truth data of a deformation, we ask the surgeon to keep steady the deforming tool and to explore around this area with the laparoscope. Under this condition, the scene remains approximately rigid. Using a calibrated laparoscope and the 5-point algorithm for RSfM [20], a 3D ground-truth point cloud representing the deformed organ is reconstructed. A set of 50 images of deformations have been used for this experiment. An average of 15 correspondences between the uterus template and the deformed images were used. They were generated using SIFT [25]. Outliers and points outside the uterus region were removed by the method proposed in [26]. In Figure 8 we show a sample of 3D reconstruction using our method from single views for the human uterus. We can see that globally our method gives meaningful 3D reconstructions according to the deformed images. Further qualitative results on other uterus tissues are reported in Figure 11. The quantitative results are shown in table I and discussed in paragraph VI-E.

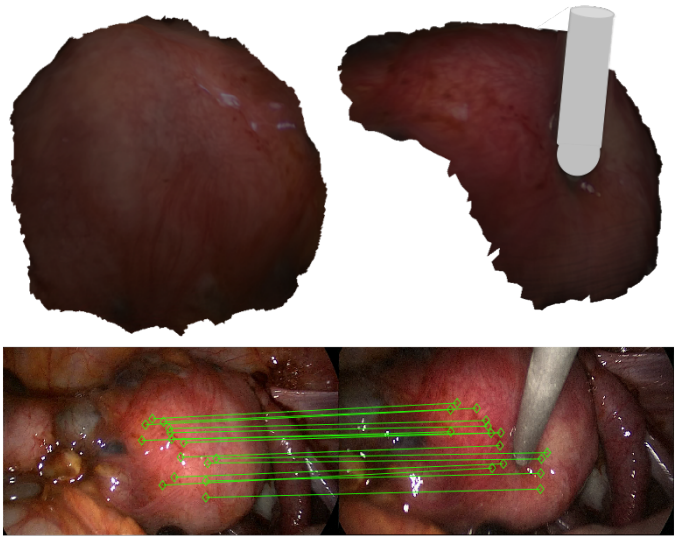

Fig. 8. Surgery of the uterus with ground-truth. Top left, the 3D template of the human uterus. The size of the box bounding the 3D shape of the uterus is $70.40 \times 65.50 \times 40.10$ mm^3. Bottom, the 15 correspondences between the template image and the deformed image. Top right, the 3D reconstructed deformed shape with the proposed method. Quantitative 3D errors of reconstruction are shown in table I.

D. Choice of the Hyper-Parameters

We computed $\lambda_1, \ldots, \lambda_6$ as described in [23]. The computed values are $\lambda^0 = (0.11,\ 0.30,\ 0.21,\ 0.21,\ 0.21,\ 0.17)$. To assess the sensitivity of the reconstruction accuracy, we evaluate the reconstruction error by varying each hyper-parameter $\lambda_i,\ i = 1 \ldots 6$, from 0.05 to 0.5 within a step of 0.01 and keeping the optimal values which provide the least 3D reconstruction error. It turns out that for deformations ranging from 2 mm to 12 mm per mesh vertex, the variation of the reconstruction error is negligible for small perturbations of the optimal values of the lambdas (see the interval of small sensitivity in Figure 10). Moreover, each hyper-parameter has an interval around the optimal value λ_i^0 where the reconstruction error is almost invariant. This observation enhances the fact that the reconstruction error is robust to small data noise. Outside the robustness intervals, we can observe that the reconstruction error is more sensitive to perturbations of the weights of the physical terms: λ_1 (conformal), λ_4 (reflectance) and λ_5 (specular), than to perturbation of the weights of the smoothing or reprojection error terms.

For this reason, the optimal λ^0s at the middle of the robust intervals give us a secure margin for reconstruction accuracy. They allow us to use the same values independently from the used data sets and for the considered range of deformation. The considered interval of deformation is completely acceptable for early surgery steps when the surgeon deforms the organ to find the locations of abnormal tissues.

E. Quantitative Evaluation and Comparison with other Methods

Table I summarizes the RMS 3D errors (c.f. appendix for computation formula) computed on both the ex-vivo and in-vivo datasets for each experimented tissue and with the maximum amount

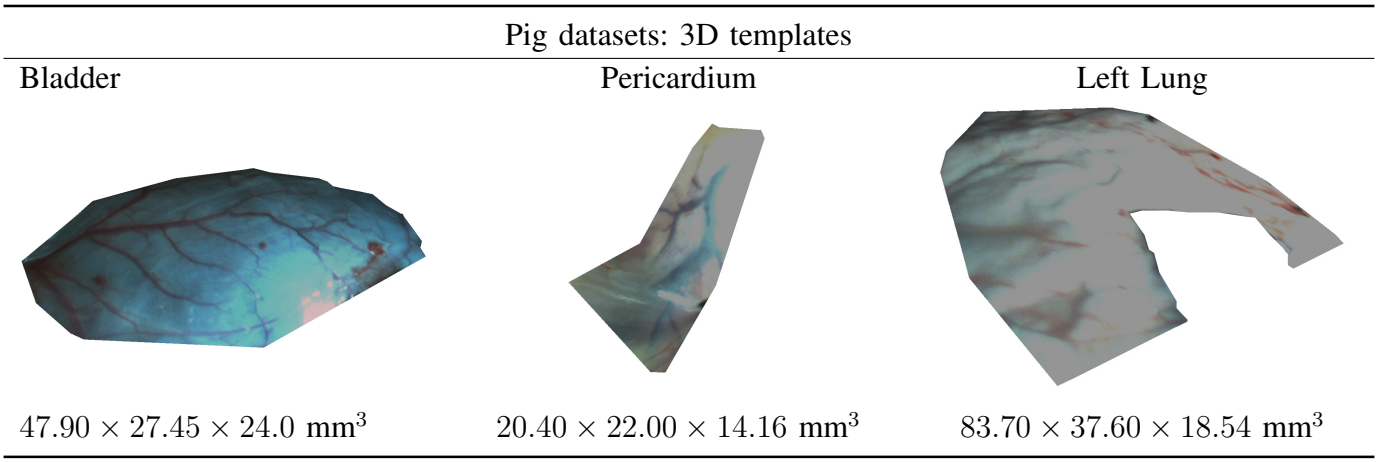

Fig. 6. Pig datasets: 3D templates of three different organ's tissues: The bladder, the pericardium and the left lung. For each template we indicate in mm the size of the box bounding the 3D shape.

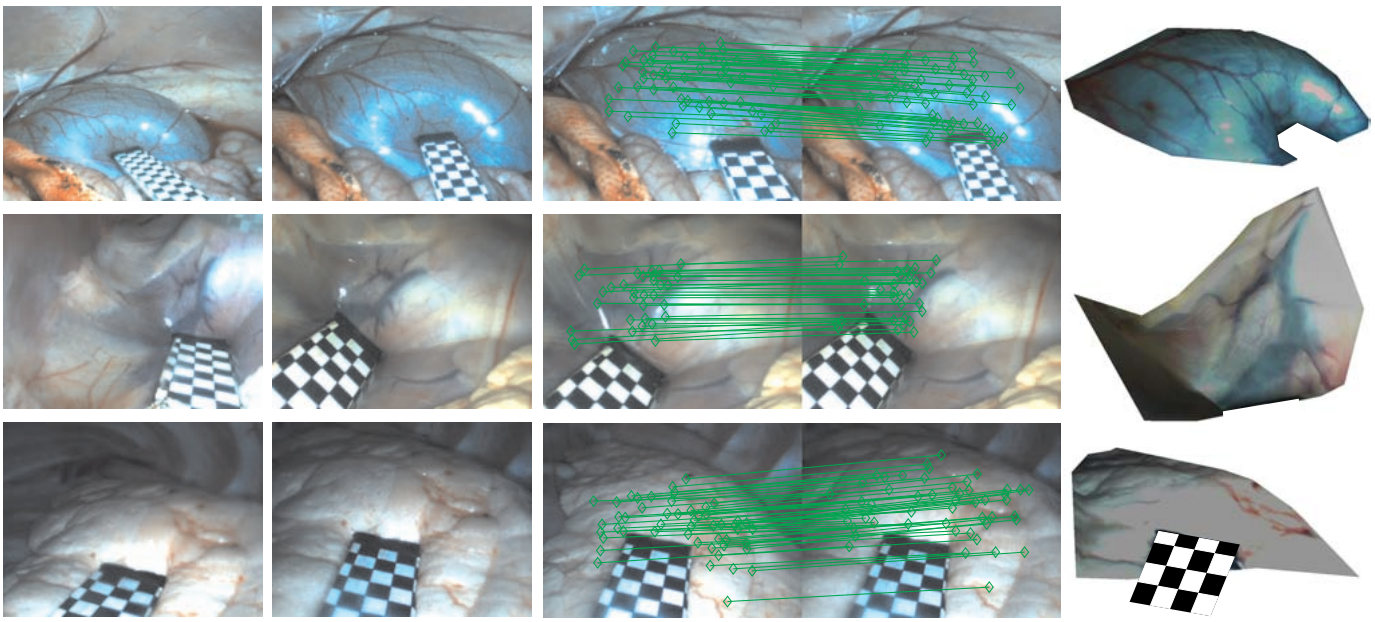

Fig. 7. In-vivo pig datasets: 3D reconstruction from a monocular laparoscope using our DSfMS method. First column: Left image from stereo view used to compute ground-truth deformation. Second column: Right image from stereo view. This image is used together with the left image to generate ground-truth 3D reconstruction. It is also used as single image to obtain 3D reconstruction with our method. Third column: correspondences between template image and right image used for the 3D reconstruction with our method. Fourth column: 3D reconstruction with our method from single image. Quantitative 3D errors of reconstruction are shown in table I.

of considered deformations (12 mm on average per mesh vertex). The proposed method **MoT-CT** provides results more accurate than the three compared methods with an average of 1.8 mm. **MoT-ON** presents an average of 2.3 mm, **MoT-LaM** presents an average of 3.7 mm and **MoT** presents an average of 5.8 mm. A detailed analysis of the results in Table I show that the **MoT-CT** method has very high performance when compared to others for moist tissues with uniform textures. This is the case of the in-vivo tissues especially the lung and the uterus. For the bladder and the pericardium tissues it has lower performance if we observe the values of the maximum errors. This is mainly due to the presence of veins at the surface where their reflectance parameters (roughness and Fresnel) are different from the rest of the tissue. The detection of such areas and the computation of the corresponding reflectance parameters can be a future development of the current approach. For the ex-vivo liver, the three algorithms have closer performance even if the best one remains **MoT-CT**. This is mainly due to the fact that the ex-vivo liver has lost its moist characteristic. The worst performance

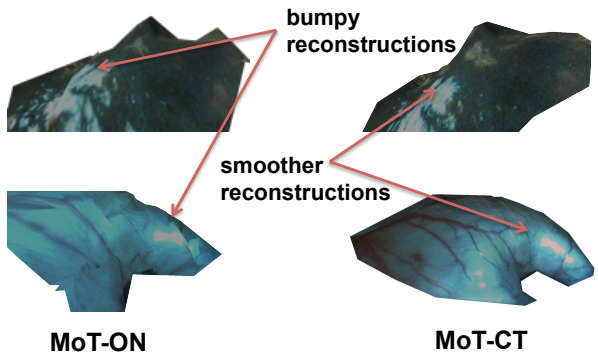

Fig. 9. Ground-Truth datasets (Lung and Bladder): Qualitative comparison of **MoT-CT** and **MoT-ON**. The former provides results smoother than the latter. The difference is more noticeable in the specular regions and their neighbourhoods.

		MoT-CT	MoT-ON	MoT-LaM
		Ex-Vivo		
Lung				
	Median	2.52	3.72	5.03
	Min	1.00	1.54	2.02
	Max	3.20	4.02	6.25
Liver				
	Median	1.22	2.49	3.35
	Min	1.03	1.54	1.70
	Max	2.11	2.42	3.75
		In-Vivo		
Bladder				
	Median	1.72	3.79	5.35
	Min	1.02	3.04	5.20
	Max	2.10	4.00	6.02
Pericardium				
	Median	1.72	3.39	4.05
	Min	1.00	2.43	3.03
	Max	3.79	4.27	6.52
Lung				
	Median	0.83	2.97	4.25
	Min	0.50	1.84	3.21
	Max	0.97	3.92	5.75
Uterus				
	Median	0.92	3.96	5.75
	Min	0.70	3.54	4.02
	Max	1.03	4.00	6.12

TABLE I
DETAILED QUANTITATIVE RESULTS FOR DIFFERENT TISSUES. THE ERRORS ARE IN MILLIMITERS.

is attributed to the ex-vivo lung because some areas have changed color and texture due to the contact with the air. In this case, as for the veins of the bladder and the pericardium, the different areas have to be identified before estimating the reflectance parameters for each one of them. This identification is not obvious and will be part of future work.

In summary, these results show that combining motion and shading cues is better than using only motion cues. Indeed, since in surgery only few correspondences can be established between the 3D template and the deformed image, it is hard for these methods to recover all the details of deformation. It appears also that the estimation of the Cook-Torrance reflectance model and its usage in shading performs better than the classic Lambertian model and the Oren-Nayar model. Indeed, the moist characteristic of the living organ's tissue is closer to the Cook-Torrance model than to a Lambertian model which assumes perfect diffuse surfaces. The assumption of the Oren-Nayar model is not sufficient to represent the specular reflectance. It represents it as Lambertian microfacets while the Cook-Torrance represents it as mirror-like microfacets. These last statements are further confirmed by Figure 9 that shows a qualitative comparison between **MoT-CT** and **MoT-ON**. In this figure we can appreciate the performance of the Cook-Torrance model in the specular regions and their neighbourhoods.

F. More Qualitative Surgery In-Vivo Data

Finally, to highlight the performance of the proposed approach on real surgery data, we use in-vivo sequences of a human uterus acquired using a monocular Karl Storz laparoscope. The frames are acquired at 30 fps and have a resolution of 1280×720. The 3D template of the uterus is generated during the laparosurgery exploration step as previously described. The Cook-Torrance parameters and the light are then calibrated as described in section IV. Complex and unpredictable deformations may occur on the uterus when the surgeon starts to examine it. A set of 500 images of deformations have been used for this experiment. An average of 25 correspondences between the uterus template and the deformed images were used. They were generated using SIFT [25]. Outliers and points outside the uterus were removed by the method proposed in [26] (Figure 11, row 2). In Figure 11, rows 3-4, we show the 3D reconstructed deformations with the corresponding deformed image in row 1. In row 4, we show synthesized views from novel camera views, and show qualitatively that the deformed uterus has been reconstructed well.

VII. Conclusion

In this paper, we presented a new method to reconstruct deforming living tissue in 3D using a single laparoscopic image and a 3D geometric and photometric template that is reconstructed in-vivo. Our 3D reconstruction pipeline DSfMS presents novel technical contributions and also a new way of tackling the 3D vision problem in laparoscopy. Ex-vivo and in-vivo experimental results show the effectiveness of combining both conformal motion cues and Cook-Torrance reflectance priors. We provided quantitative comparison with other methods which combine motion cues with a Lambertian or an Oren-Nayar reflectance models.

We showed that for moist tissue with the same reflectance property, the Cook-Torrance model is the best candidate. Future developments of our work will be focused on the detection of areas with different reflectance parameters.

Appendix

Error Measurement

In order to evaluate the performance of our approach, we computed the RMS 3D error in mm as:

$$\sqrt{\frac{\sum_{i=1}^{N} \parallel \mathbf{v}'_i - \mathbf{v}_i \parallel^2}{N}},$$

where $\{\mathbf{v}'_i\}_{i=1,\dots,N}$ are the vertices of the 3D reconstructed mesh and $\{\mathbf{v}_i\}_{i=1,\dots,N}$ are the vertices of the ground truth mesh.

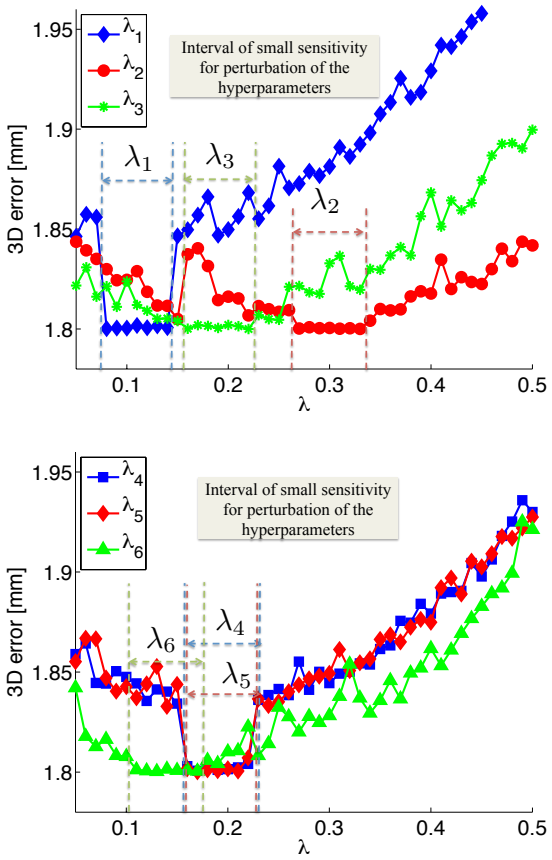

Fig. 10. Sensitivity of the reconstruction error with respect to the hyper-parameters. Each computed hyper-parameter has an interval of confidence in which the reconstruction error is robust to small perturbations. Outside this interval the error grows slightly to reach 1.96 mm at most (we recall that the minimal 3D error is 1.8 mm on average with our method). This quantitative study shows that the computed hyper-parameters are optimal and robust to small perturbations.

References

[1] F. Moreno-Noguer, M. Salzmann, V. Lepetit, and P. Fua, "Capturing 3D stretchable surfaces from single images in closed form," in *CVPR*, 2009.

[2] M. Salzmann, R. Urtasun, and P. Fua, "Local deformation models for monocular 3D shape recovery," in *CVPR*, 2008.

[3] F. Brunet, R. Hartley, A. Bartoli, N. Navab, and R. Malgouyres, "Monocular template-based reconstruction of smooth and inextensible surfaces," in *ACCV*, 2010.

[4] A. Malti, A. Bartoli, and T. Collins, "Template-based conformal shape-from-motion from registered laparoscopic images," in *MIUA*, 2011.

[5] D. Stoyanov, M. V. Scarzanella, P. Pratt, and G.-Z. Yang, "Real-time stereo reconstruction in robotically assisted minimally invasive surgery," in *MICCAI*, 2011.

[6] E. Prados, F. Camilli, and O. Faugeras, "A unifying and rigorous shape from shading method adapted to realistic data and applications," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 3, pp. 307–328, 2006.

[7] A. Malti, A. Bartoli, and T. Collins, "Template-based conformal shape-from-motion-and-shading for laparoscopy," in *IPCAI*, 2012.

[8] R. L. Cook and K. E. Torrance, "A reflectance model for computer graphics," in *SIGGRAPH*, 1981.

[9] L. Wolff, S. Nayar, and M. Oren, "Improved diffuse reflection models for computer vision," *International Journal of Computer Vision*, vol. 30, no. 1, pp. 55–71, 1998.

[10] L. Maier-Hein, P. Mountney, A. Bartoli, H. Elhawary, D. Elson, A. Groch, A. Kolb, M. Rodrigues, J. Sorger, S. Speidel, and D. Stoyanov, "Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery," *Medical Image Analysis*, 17(8):974-996, 2013.

[11] M. Hayashibe, N. Suzuki, and Y. Nakamura, "Intraoperative fast 3d shape recovery of abdominal organs in laparoscopy," in *MICCAI*, 2002.

[12] ——, "Laser-scan endoscope system for intraoperativegeometry acquisition and surgical robot safety management." *MIA*, vol. 10, pp. 509–519, 2006.

[13] J. Penne, K. Holler, M. Sturmer, T. Schrauder, A. Schneider, R. Engelbrecht, H. Feubner, B. Schmauss, and J. Hornegger, "Time-of-fligth 3-d endoscopy," in *MICCAI*, 2009.

[14] G. Hager, B. Vagvolgyi, and D. Yuh, "Stereoscopic video overlay with deformable registration," in *Medicine Meets Virtual Reality*, 2007.

[15] J. Totz, P. Mountney, , and D. S. G. Yang, "Dense surface reconstruction for enhanced navigation in MIS," in *MICCAI*, 2011, pp. 89–96.

[16] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah, "Shape from

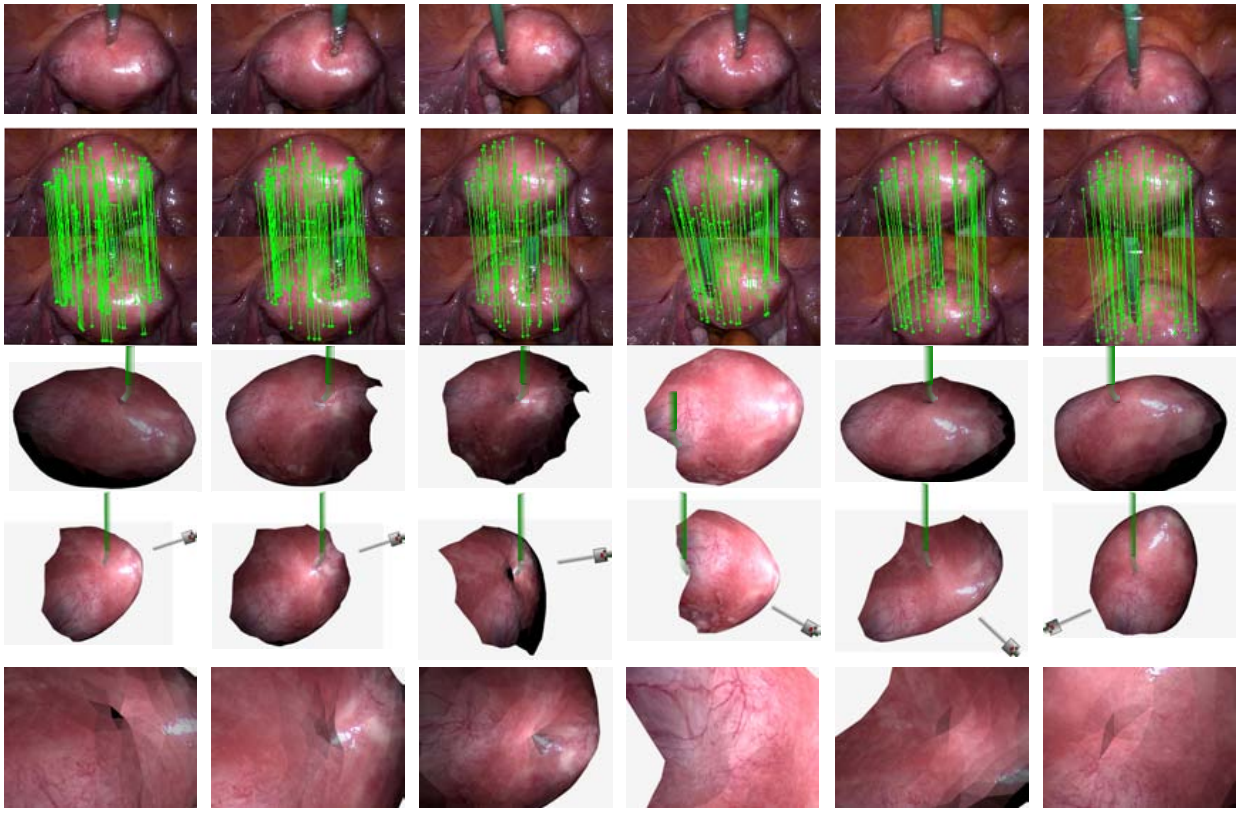

Fig. 11. 3D reconstruction on an in-vivo video sequence from a monocular laparoscope using our conformal method. First row: Single 2D views of uterus deformation with a surgery tool. Second row: Point correspondences between the template and deformed images. Third row: 3D reconstruction using our DSfMS method. Each 3D reconstruction is done using the single view above. The view is given in the laparoscope's view point. Fourth row: 3D deformed surface seen from a different point of view which provides visualization of the self-occluded part. Fifth row: Zoom in the deformed area.

shading: A survey," *PAMI*, vol. 21, no. 8, pp. 690–706, 1999.

[17] L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz, "Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo," in *ICCV*, 2003.

[18] C. Wu, K. Varanasi, Y. Liu, H.-P. Seidel, and C. Theobalt, "Shading-based dynamic shape refinement from multi-view video under general illumination," in *ICCV*, 2011.

[19] A. Malti and A. Bartoli, "Estimating the Cook-Torrance BRDF parameters in-vivo from laparoscopic images," in *Workshop Augmented Environement at MICCAI*, 2012.

[20] D. Nistèr, "An efficient solution to the five-point relative pose problem," IEEE *Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 756–770, 2004.

[21] G. Zou, J. Hu, X. Gu, and J. Hua, "Area-preserving surface flattening using lie advection," in *MICCAI*, 2011.

[22] D. Yoo, "Three-dimensional morphing of similar shapes using a template mesh," *International Journal of Precision Engineering and Manufacturing*, 2009.

[23] B. Compte, A. Bartoli, and D. Pizarro, "Constant flow sampling: A method to automatically select the regularization parameter in image registration," in *Fifth Workshop on Biomedical Image Registration*, 2012.

[24] A. Chhatkuli, A. Bartoli, A. Malti, and T. Collins, "Live image parsing in uterine laparoscopy," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2014.

[25] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[26] P. F. Alcantarilla and A. Bartoli, "Deformable 3D reconstruction with an object database," in *BMVC*, 2012.