LIVE IMAGE PARSING IN UTERINE LAPAROSCOPY

Ajad Chhatkuli, Adrien Bartoli, Abed Malti, Toby Collins

ISIT, UMR 6284 CNRS/Université d'Auvergne, Clermont-Ferrand, France

ABSTRACT

Augmented Reality (AR) can improve the information delivery to surgeons. In laparosurgery, the primary goal of AR is to provide multimodal information overlaid in live laparoscopic videos. For gynecologic laparoscopy, the 3D reconstruction of uterus and its deformable registration to preoperative data form the major problems in AR. Shape-from-Shading (SfS) and inter-frame registration require an accurate identification of the uterus region, the occlusions due to surgical tools, specularities, and other tissues. We propose a cascaded patient-specific real-time segmentation method to identify these four important regions. We use a color based Gaussian Mixture Model (GMM) to segment the tools and a more elaborate color and texture model to segment the uterus. The specularities are obtained by a saturation test. We show that our segmentation improves SfS and inter-frame registration of the uterus.

***Index Terms*—** Segmentation, laparoscopy, uterus.

1. INTRODUCTION

AR in gynecologic laparoscopy is expected to be of practical importance for several pathologies [2]. Implementing AR to overlay preoperative data such as MRI extracted tumors into the live laparoscopic video is, nonetheless, a colossal, difficult and still unsolved task. Live image parsing is one of the main components of AR. By live image parsing implies the segmentation of laparoscopic video frames into four major classes: uterus, tool, specularity and other tissues as shown in Figure 1.

Identifying the uterus and tools has a particular importance in 3D reconstruction required for AR. Several advances have been made in recent years in monocular 3D reconstruction with laparoscopic images, using for instance a combination of SfS and Shape-from-Motion (SfM) [9]. However, in many situations related to pathologies of the uterus, the deformation defeats SfM methods. On the other hand, SfS does not suffer from the deformations [4]. SfS benefits from image parsing: the surface normal can be fixed in specular segments, and the 3D reconstruction domain is limited to the uterus for many pathologies. Live segmentation also helps inter-frame registration by eliminating a large amount of false matches between the video frames. As in the case of SfS, it also limits the image domain, potentially providing an overall speed-up. AR also directly benefits from image parsing as overlays should not erase the tools.

We present a carefully engineered cascaded system of multiple processes to obtain the live parsing of uterine laparoscopic images. The tools in laparoscopy have discriminant color. As such, we exploit the color information with a Gaussian Mixture Model (GMM) to learn the tool characteristics and use likelihood estimation and graph cut to segment input frames. For the uterus, however, color alone was found to be insufficient to arrive at a satisfactory segmentation performance. Thus we use Support Vector Machines (SVM) to learn dense feature descriptors that encode both color and texture in a patient-specific manner. Finally, we also segment the specularities in the uterus to obtain a four class segmentation result.

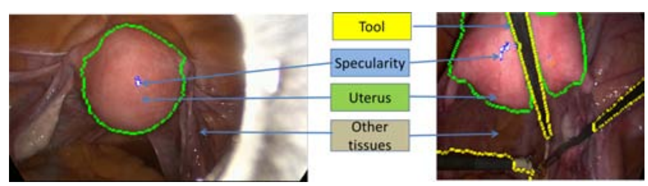

Fig. 1: Example outputs of our live parsing with labels.

We validate the impact of parsing in two ways. In the first component of 3D reconstruction by SfS, segmentation is used to exclude tools and other tissues to obtain a reconstruction with less artifacts. In addition, specularities are used by SfS to get the surface normals at those locations. It is to be noted that such a segmentation would also be crucial in other methods of 3D reconstruction while the uterus is deforming [5]. Similarly, we exploit segmentation in inter-frame registration to reduce the amount of false putative keypoint matches.

2. SEGMENTATION

We achieve parsing by two main binary segmentation processes and some post-processing. In order to perform each segmentation process we use specific constraints and methods optimized to the particular segmentation task. A process diagram is shown in Figure 2.

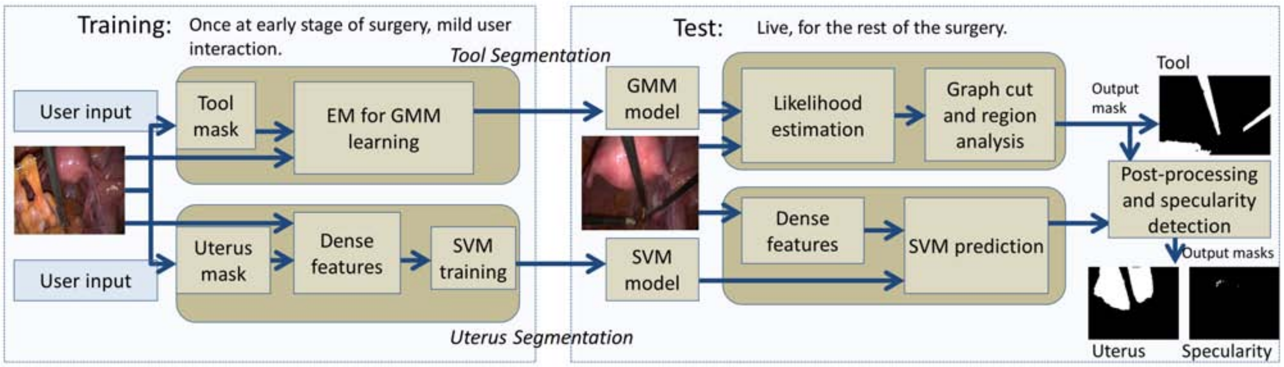

Fig. 2: A detailed process diagram of our four-class parsing. The actual processing is carried out after downsampling the input images for speed.

2.1. Tool Segmentation from Color

One way to obtain tool segmentation is to use supervised classification of pixels with a feature descriptor based on gradient and color [1]. Another class of methods uses tracking algorithm with geometric constraints [10]. We present an alternative approach that works at frame rate using GMMs. A GMM can be used to encode the first order and second order moments of color in images. Tools in laparoscopy have specific colors that are quite different from the rest of the image. It motivates the use of color GMMs of the tool for our task. A minimal interaction with the user is also preferred for the automatic segmentation procedure. The GMMs of the tool is learned from a single image, where the user draws a polygonal region to extract the tool pixels. The whole process is done in normalized RGB in order to handle illumination changes.

A GMM can be represented by:

$$p(x|\lambda) = \sum_{i=1}^{K} w_i g(x|\mu_i, \Sigma_i).$$

Here, $K = 4$ is a hand-tuned parameter in our case, chosen in order to represent each pixel color with 4 Gaussian components. The Gaussian parameters are represented by $\lambda = \{w_i, \mu_i, \Sigma_i\}$, with $i = 1, \ldots, K$. The parameters are estimated with the popular Expectation Maximization (EM) algorithm.

The second step in the tool segmentation process is to construct a probability map for test-image pixels using the trained GMMs. With likelihood estimation, a probability map is obtained that indicates the probability of an image pixel belonging to the tool. The result of the probability map is then fed to a Graph Cut algorithm to obtain a smooth estimate of tool segmentation. Further post-processing based on connected region analysis then eliminates regions that do not touch an image boundary or that are too small (< 50 pixels). The process is done at an image resolution of 350×250 or 350×200 depending on the original aspect ratio of the input image.

2.2. Uterus Segmentation from Color and Texture

Organ segmentation in laparoscopic images is a very challenging task and the literature regarding such is rare. [8] presents an unsupervised algorithm that lacks semantics and was not found to be suitable for uterus segmentation. The uterus segmentation task is further complicated by inter-patient variabilities such as color and texture and intra-patient variabilities such as the viewpoint, scale and the associated changes in texture perception. Our approach to segmentation is to proceed with supervised learning of a patient-specific uterus model with dense feature descriptors that exploit both color and texture.

AM-FM texture descriptor. Methods based on AM-FM image analysis represent an image as a combination of amplitude and frequency modulated signals [6]. To capture cues from the uterus and off the uterus (including other organs, blood vessels, etc), we use Gabor filters in AM-FM image analysis as in [6]. An AM-FM model can describe an intensity image $f(x, y)$ with spatial coordinates (x, y) in the form of the equation:

$$f(x, y) = a(x, y) \cos(\phi(x, y)), \quad (1)$$

where the local image contrast is represented by the amplitude signal a and the image structure by the instantaneous frequency vector in terms of phase ϕ:

$$\vec{\omega}(x, y) = \nabla\phi(x, y) = \left(\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}\right)(x, y). \quad (2)$$

The parameters a and $\vec{\omega}$ are estimated with an Energy Separation Algorithm (ESA) [6].

To have a better characterization of texture and frequency information, ESA has been applied on separate Gabor filtered channels [6]. This also regularizes the parameter estimation by replacing image derivatives with filter derivatives. To capture different visual cues, we use two different sets of Gabor filter frequencies as in [6], but with different parameters. The first set that gives high responses on contours and high frequency variations off the uterus has center frequencies in the range of 0.09 to 0.4 radians per pixel. The second set of filters that gives high responses on smooth regions such as the uterus with very small spatial changes has center frequencies in the range of 0.02 to 0.08 radians per pixel. 40 different filters from each set were used to get a descriptor to model the uterus region in the luminance channel of CIE $L^*a^*b^*$ color space.

Texture descriptor formulation. ESA gives the amplitude and frequency quantities for each pixel of the image after demodulation. The descriptor thus consists of a signal magnitude and a frequency magnitude from each of the 80 different filters. Consequently, the dimension of the descriptor becomes very large to up to Image_size $\times 2 \times 2 \times 40$. Using such high dimensional features directly for classification was found to result in poor performance. Thus we reduced the dimensions of the descriptor to Image_size $\times 2 \times 2$ by taking only the maximum value of the component for each set of filters. This simple form of dimension reduction was preferred over Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) for its speed and also because the output performance due to dominant components was in par with the PCA and LDA methods.

Color descriptor. In laparoscopic images, color is probably the most important cue that can discriminate uterus from the rest of the tissues and even tools. However, the RGB representation of color does not capture the similarity of uterus color in different views. To obtain a color descriptor that is discriminant, while retaining similar properties in illumination varying situations is an open field of its own. We choose the classical CIE $L^*a^*b^*$ color space to extract color information using the a^* and b^* channels, while the L^* channel is used by the AM-FM descriptor formulation. This approach of combining the texture information with color was found to be simple as well as effective, hence suitable for an actual surgical scenario.

Classification and post-processing. The feature descriptors are obtained by combining the 4-dimensional AM-FM descriptors and the 2-dimensional color descriptor for each pixel. To classify the pixels into uterus and non-uterus classes, we use a linear SVM from the LIBSVM library [3], which was found to have sufficient accuracy and very low training and prediction time. The pixel-wise binary classification was found to give largely an accurate segmentation. To remove small falsely detected regions outside the uterus and to eliminate small false negative regions we use connected region analysis after a smoothing operation. The connected regions other than the largest two and those with the major axis length more than 2.5 times that of the minor axis length were eliminated. In addition, the result of tool segmentation was used to remove the false detection of tool pixels as uterus. We also use the connected regions to fill holes (such as those left by specularities) in the segmentation mask. The result of the whole process is a mask for the uterus region.

2.3. Specularity Segmentation from Saturation

The specularities are found by thresholding on luminance (0.95) and saturation (0.9) channels in the HSV color space after a Gaussian smoothing. The saturation and luminance test together select only white pixels with sufficient luminance so that they can be safely assumed as specularities. The binary segmentation obtained from the process is then combined with the uterus mask to get the specularities in uterus. The final result is the complete 4-class parsing of a laparoscopic video frame.

3. RESULTS AND APPLICATIONS

Our implementation was done in both C++ and MATLAB. The MATLAB implementation was used to evaluate our methods, to choose their parameters and to compare them against other approaches. The optimal design was implemented in C++ using the OpenCV and CUDA libraries, to test for the real-time feasibility of our system. The uterus segmentation forms the most critical part of the system and therefore its detailed analysis was done independently in a well structured dataset. The dataset contains laparoscopic videos for 15 different patients, each with images that were split into training, validation and test set. The number of images used was 3 in the training set and 5 in the others for each patient. The images were chosen to include difficult cases and views with large variabilities from the training images. We obtained a Dice similarity ratio of 80.44% and a false detection rate of 3.29% using 2 training images for each patient. Most inaccuracies in the results were due to undersegmentation of the uterus, which is natural around strong contours due to the use of filtering based descriptors. Moreover, for a typical surgical scenario, where the camera is focused on the uterus, the accuracy was observed to be much higher. Tool segmentation showed good performance against several videos, while giving highest accuracies in presence of green colored tools. The OpenCV implementation of the complete system takes around 250ms to segment each frame at a resolution of 350×200 in a Dell Alienware desktop PC of 2011. Figure 3 shows some results of the live parsing.

To validate the motivation for segmentation we experimented 3D reconstruction with and without using segmentation. We used perspective SfS [7] to reconstruct the 3D shape of the segmented uterus. The results in Figure 4 show better performance with segmentation around the discontinuities (tools). The results clearly imply the impossibility of

Fig. 3: Representative outputs of segmentation (see Figure 1 for legend).

using the reconstruction with unsegmented image to perform further tasks such as 3D matching. Similarly, the results of matching using SIFT in different pair of frames are shown in Figure 5 with and without segmentation. The results show correspondences between two images of a uterus, where correspondences outside the uterus are eliminated. It is what we would require for a deformable registration between two views of a deforming uterus. One critical part in such applications is the elimination of specular regions that would otherwise lead to false matches, thus supporting the need for a four-class live parsing.

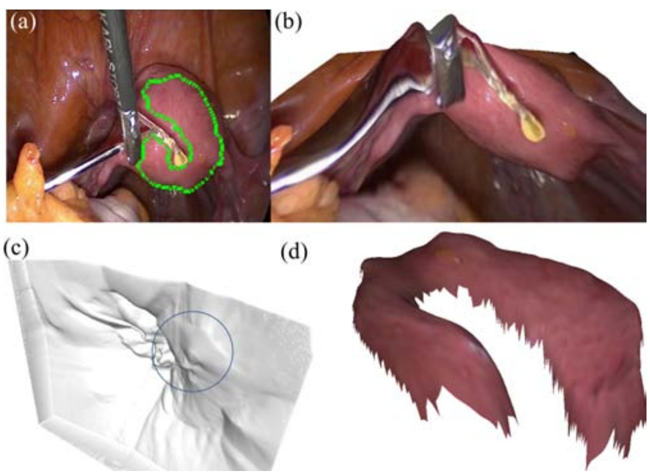

Fig. 4: (a) Input image (b) SfS result without segmentation: texture mapped view (c) second (surface) view with a circle indicating the region of interest (d) SfS result with segmentation.

4. CONCLUSIONS

We have presented a cascaded method for live parsing of uterine images in laparoscopic uterus surgery. Our results show that such segmentation has several applications, particularly towards aiding the components of AR. Besides SfS and registration, numerous other applications may exist. In order to improve the current system, methods based on Conditional Random Fields could be explored. We are currently working on fitting our live parsing into a complete AR suite; this may impact medical computer vision importantly.

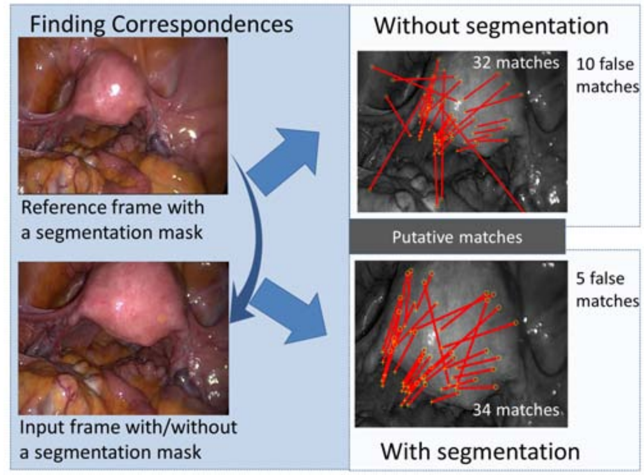

Fig. 5: Results of matching between a reference and an input image.

5. REFERENCES

[1] M. Allan, S. Ourselin, S. Thompson, D. J. Hawkes, J. Kelly, and D. Stoyanov. Toward detection and localization of instruments in minimally invasive surgery. *IEEE Trans. Biomed. Engineering*, 60(4):1050–1058, 2013.

[2] 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*, 79(6):858–863, 2012.

[3] C. Chang and C. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):1–27, 2011.

[4] T. Collins and A. Bartoli. Towards live monocular 3D laparoscopy using shading and specularity information. In *IPCAI*, 2012.

[5] T. Collins, D. Pizarro, A. Bartoli, M. Canis, and N. Bourdel. Realtime wide-baseline registration of the uterus in laparoscopic videos using multiple texture maps. In *AE-CAI*, 2013.

[6] I. Kokkinos, G. Evangelopoulos, and P. Maragos. Texture analysis and segmentation using modulation features, generative models, and weighted curve evolution. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(1):142–157, 2009.

[7] M. V. Scarzanella, D. Stoyanov, and G.-Z. Yang. Metric depth recovery from monocular images using shape-from-shading and specularities. In *ICIP*, 2012.

[8] Y. Shu, G. Bilodeau, and F. Cheriet. Segmentation of laparoscopic images: Integrating graph-based segmentation and multistage region merging. In *CRV*, 2005.

[9] H. N. Tokgozoglu, E. M. Meisner, M. Kazhdan, and G. D. Hager. Color-based hybrid reconstruction for endoscopy. In *Workshops on Medical Computer Vision at CVPR*, 2012.

[10] R. Wolf, J. Duchateau, P. Cinquin, and S. Voros. 3D tracking of laparoscopic instruments using statistical and geometric modeling. In *MICCAI*, 2011.