

# Vision Based Robot Assistance in TTTS Fetal Surgery

Narcís Sayols<sup>1,\*</sup>, Albert Hernansanz<sup>1,\*</sup>, Johanna Parra<sup>2,\*</sup>, Elisenda Eixarch<sup>2,3</sup>, Eduard Gratacós<sup>2,3</sup>, Josep Amat<sup>1</sup> and Alicia Casals<sup>1</sup>

**Abstract**—This paper presents an accurate and robust tracking vision algorithm for Fetoscopic Laser Photo-coagulation (FLP) surgery for Twin-Twin Transfusion Syndrome (TTTS). The aim of the proposed method is to assist surgeons during anastomosis localization, coagulation and review using a tele-operated robotic system. The algorithm computes the relative position of the fetoscope tool tip with respect to the placenta, via local vascular structure registration.

The algorithm uses image features (local superficial vascular structures of the placenta’s surface) to automatically match consecutive fetoscopic images. It is composed of three sequential steps: image processing (filtering, binarization and vascular structures segmentation); relevant Points Of Interest (POIs) selection; and image registration between consecutive images.

The algorithm has to deal with the low quality of fetoscopic images, the liquid and dirty environment inside the placenta jointly with the thin diameter of the fetoscope optics and low amount of environment light reduces the image quality. The obtained images are blurred, noisy and with very poor color components.

The tracking system has been tested using real video sequences of FLP surgery for TTTS. The computational performance enables real time tracking, locally guiding the robot over the placenta’s surface with enough accuracy.

## I. INTRODUCTION

Fetoscopy has been introduced in fetal diagnosis and therapy and its application in some cases affecting the fetus is the only treatment that can be offered. Laser coagulation of placental vessels for Twin-Twin transfusion syndrome (TTTS) is the most common surgery carried out today[1], which is associated with survival rates of at least one twin ranging from 75 to 85%[2]. This syndrome is a severe complication that affects about 10-15% of monochorionic twin pregnancies[3] and with expectant management, is associated with a mortality rate of about 80-100%. TTTS is the result from an unbalanced intertwin blood flow from the donor to the recipient twin through placental vascular anastomoses.

<sup>1</sup>Narcís Sayols, Albert Hernansanz, Josep Amat and Alicia Casals are with the Center of Research in Biomedical Engineering, Universitat Politècnica de Catalunya, Jordi Girona 1-3, Barcelona, Spain. {narcis.sayols, albert.hernansanz, josep.amat, alicia.casals}@upc.edu

<sup>2</sup>Johanna Parra, Elisenda Eixarch and Eduard Gratacós are with Fetal i+D Fetal Medicine Research Center, BCNatal - Barcelona Center for Maternal- Fetal and Neonatal Medicine (Hospital Clínic and Hospital Sant Joan de Déu), Institut Clínic de Ginecologia, Obstetrícia i Neonatologia, Institut d’Investigacions Biomèdiques August Pi i Sunyer, Universitat de Barcelona, Spain. jparra@sjdhospitalbarcelona.org, {eixarch, gratacos}@clinic.cat

<sup>3</sup>Elisenda Eixarch and Eduard Gratacós are with Centre for Biomedical Research on Rare Diseases (CIBER-ER), Barcelona, Spain

\*Authors have equally contributed to this work

The first-line treatment of severe TTTS is fetoscopic laser photocoagulation of those anastomoses.

Fetoscopic laser surgery uses intrauterine fetoscopy with a laser fibre. A 2 to 3mm cannula is percutaneously inserted into the recipient sac under ultrasound guidance. A rigid fetoscope with a straight sheath is used for a posterior placenta, and a fetoscope with a curved sheath is used for an anterior placenta. A 600- $\mu$ m laser fibre is passed through the operative channel of the sheath of the fetoscope, and a (Nd:YAG) or diode (semi-conductor) laser is used. All communicating vessels between the twins on the chorionic plate of the placenta (including arteriovenous [AV], arterioarterial [AA] and venovenous [VV] anastomoses) are coagulated with a nontouch technique. The goal of surgery is to ablate all intertwine anastomoses[4].

Actual FLP technique can be improved in several aspects: time reduction (anaesthesia exposition of foetuses), anastomoses ablation accuracy (reducing the affected vessels and placenta’s ablated surface) and increase patients safety avoiding fetoscope contact with the placenta due to surgical environment movements (breathing, aorta pulsation in posterior placenta and, occasionally, spasmodic and non-predictable movements, e.g. coughing). The surgical procedure consists of a sequence of steps including: localization of umbilical insertions, search of the equator (middle region between umbilical insertions where the anastomoses are located with higher probability), search for anastomoses, planning the order of their coagulation, navigate to anastomoses and their coagulation and, finally, a review and, if required, re-coagulation of each anastomose.

The proposed solution is based on the analysis of the fetoscopic images to extract relevant information (vascular structures) and infer the relative position between fetoscope and placenta. This information can be used by the robotic system to stabilize and perform fine tracking of local Regions of Interest (ROIs). The introduction of a teleoperated system with pseudo-autonomous fetoscope guidance can improve actual FLP performance. Due to the nature of the surgery, with deformable tissues and non-predictable movements, pre-planning of trajectories is not possible. A useful solution is to provide the system with the relative position of the fetoscope with respect to the placenta and associate the acquired images with these positions. This information enables the possibility of real-time collision risk analysis, image stabilization and fine tracking of placenta’s Points of Interest (POIs).

This paper is structured as follows: the following section reviews the state of art of new FLP technological solutions

and automatic vessel detection techniques. Then, Section III describes the visual tracking algorithm and shows the obtained results using real video sequences of FLP surgeries. Section IV introduces the teleoperation system where the algorithm is integrated. Conclusions and next research steps are described in the last section.

## II. RELATED WORK

Fetoscopy has been object of study to obtain both, the placenta's 3D vascular reconstruction and its surface mosaicking to provide the surgeons with pre-operative or intra-operative information. Several works generate a 3D reconstruction based on pre-operative medical images. In this respect, in [5] MRI images are used to obtain primary data, while in [6] 3D power angiography and color imaging Dopplers are used to obtain blood flow. In [7] a 3D reconstruction of arterial network from micro-computed tomography is obtained for the study of fluid dynamics of the placenta (experimentally, from a rat). All these 3D models can be used to detect and reconstruct structures, identify branches and anastomoses, and generate a knowledge dataset of placental vascular structures. In [8], the generation of a vessels map using MR angiography reduces the computation complexity (from a 3D analysis to 2D continuity analysis plus cross section detection). In [9], the objective is the tracking of capillary vessels using a pseudo-autonomous off-line method, selecting seeds manually. The authors use the formulation of graph construction for their reconstruction approach, using different classes (curves, branches, half lines, etc.). The method is designed to detect anomalies in capillary structures. A similar problem is treated in [10], detecting retina's vessels using the global image. The method is based on the automatic detection of seed points and the reconstruction of the vessels structures using the local gradient information. These last algorithms are used for non real-time medical analysis.

Some works generate a complete placenta's surface mosaic to provide the surgeons with a map of the superficial vascular structures for TTTS surgery. In [11] the problem is addressed with an automatic detection of POIs (blobs) and associating a characteristics vector to each of them. To match images and compute their respective transformations, the proposed algorithm uses the maximal vector similitudes. Finally, the mosaic is refined with the sum of the squared differences as the optimization norm, minimizing transitions between paired images. This method generates noticeable results (back projection error of 0.6 pixel), but cannot be applied to real time tracking: it needs 3.4sec for every pair of images. In [12], the authors apply a novel approach based on the use of convolutional neural networks, with the need of a first learning phase. The used images come from a simulated placenta, with better visual conditions than in real scenarios. Following with placenta's reconstruction, [13] generates the mosaic combining image registration with the Square Root Unscented Kalman Filter to use the temporal information and an electromagnetic tracker system to reduce the accumulative error, thus obtaining the absolute position of the fetoscope. The resulting system is tested on an ex-vivo placenta in a

free and clean environment. There is no information on the required time to match pairs of images. Finally, in [14] the Scale Invariant Feature Transform (SIFT) method is used to generate the mosaic. To reduce the required computational time, the method uses GPUs parallel computation. The initial results are promising in terms of coherence on the obtained vascular structures, but it is tested on an ex-vivo paper printed placenta model, avoiding environmental difficulties.

Neural networks are used in [15] to classify the fetoscopic images into three different actions: ablation, targeting and others. The objective is to obtain automatically a workflow of the fetoscopic surgery procedure. The authors empathise on the high variability and poor quality of the obtained images. To overcome with this limitation and determine which action is being performed, they use convolutional neural networks. The color of laser light is one of the main image features used to classify the surgical action.

The presented method offers a valid solution due to the obtained results with real images. This algorithm presents robustness in front of image variability (high differences between each patients placenta surface, environmental illumination, dirtiness, etc.) and poor image quality. A second benefit is the computational efficiency and the possibility of parallelization over GPUs.

## III. IMAGE ACQUISITION IN FETOSCOPY

In fetoscopy, image processing constitutes a complex problem due to the low quality of the obtained images. Several reasons cause the acquired images to turn out to be of very poor quality. The first problem that arises in fetoscopy image acquisition is the optics diameter, which is  $2 \sim 3mm\varnothing$  and illumination of the space is low. Consequently, the amount of light is limited, resulting into a noisy image. The required camera's long exposition time jointly with the camera and placenta's movements generate blurred images. A second problem produced by the optics thickness is the visualization of the optical fibers into the image in a similar form of a non structured honeycomb. It is difficult to extract a filtering mask because of the lack of a well defined structure and the free rotation between the fetoscope (containing the fibers) and the camera sensor. In addition, the environment and the placenta also interfere with the correct image acquisition. Placenta's textures present a high variability, preventing the extraction of a defined pattern from the images. The color of the blood vessels is not homogeneous, hindering the extraction of structures information and their continuity. The liquid environment, amniotic fluid, deteriorates the obtained images. The light spot of the endoscope, when is close to the placenta, saturates partially or totally the image. The floating detritus and the air bubbles also saturate the image. Both of them appear as moving big spots with unpredictable trajectories.

## IV. VISUAL TRACKING ALGORITHM

The pseudo-autonomous fetoscope guidance needs its relative position with respect to the placenta. The fetoscope pose is continuously provided by the robotic arm while

the visual tracking algorithm computes the transformation between two consecutive images. Once obtained the images transformation and known the fetoscope pose, it is possible to control the relative transformation between fetoscope and the focused local placenta region. This transformation opens the possibility of an active camera tracking of an interesting point (e.g. an anastomoses) and maintain a fixed distance between the fetoscope and the placenta.

The developed visual algorithm uses the vascular structures of the placenta surface to extract relevant points and compute the transformation between two consecutive images.

The algorithm is divided into three sequential steps: first, filtering, binarization and segmentation of the new image, to extract the vessels structures. Second, obtaining the image POIs. These POIs are used in the third step to compute the local registration in the  $XY$  plane and, finally, computing the global transformation.

#### A. Image Binarization and Segmentation

The segmentation of the different arteries and vessels of the placenta relies on a complex process due to poor quality of the fetal images. A block diagram of the segmentation and tracking process is presented in Fig. 1.

First, given a fetoscopy image, Fig. 2.a, the algorithm extracts the green channel of the images because this channel has a better contrast between the vascular structures and the background according to [16], [17] and applies a median blur filter to remove noise, Fig. 2.b. Second, the algorithm computes a Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance the contrast between the different structures in the image, Fig. 2.c. Third, an adaptive thresholding binarization is applied using the algorithm explained in [18], Fig. 2.d. Finally, a serie of postprocessing filters is applied, a speckles filter to eliminate the small components of the binary image, Fig. 2.e, and a morphological closing filter to fill the small holes in the detected structures, Fig 2.f.

#### B. Points of Interest Selection

In order for the algorithm to reach a solution, a set of suitable POIs are needed. These POIs must be distributed uniformly along the image as much as possible to obtain the local translation of all the zones to calculate the global transformation. Moreover, these POIs, in the binary segmentation, must also fulfill that the window around them has a homogeneous distribution of black and white pixels to obtain a good displacement in the local correlation step.

The strategy used to select the POIs is as follows. Let  $B_i$  be the binary segmented image of  $I_i$ . Then, the algorithm searches for a set of circles (two for the presented experiments) with centre in  $C_i$  and takes as point candidates those corresponding to the intersection of these circles with the contour of the segmented areas. These circles are updated at each image by using those in the previous image and varying their radius values until a minimum of intersections with the segmented areas is reached, and also looking for the POIs to be uniformly distributed in all directions with respect to the centre of the image, see Fig. 3.

Let  $\{P_1, \dots, P_{m_i}\}$  be the candidate POIs for image  $B_i$  and  $Q = \{Q_1, \dots, Q_{m_{i-1}}\}$  the POIs obtained using the new circles in image  $B_{i-1}$ . Then,  $\forall P_j$  the algorithm searches if a correspondence exists in  $\{Q_1, \dots, Q_{m_{i-1}}\}$ , see Fig. 4. If this correspondence does not exist, the algorithm removes the corresponding  $P_j$  points from the set. If two POIs are too close, the algorithm takes the mean of both POIs as a new point. Once all the POIs have been tested, if there are not enough POIs, the algorithm searches again different circles that verify all the previous conditions. In the experiments, a minimum of six POIs are required to obtain a robust solution.

#### C. Image Registration Algorithm

This subsection explains the algorithm used to find the transformation that matches two consecutive fetoscopic images. The algorithm uses an Eucliden transformation with 6 degrees of freedom (3 for position and 3 for orientation). A block diagram with all the algorithms' steps to find the transformation between the images is shown in Fig. 1. The idea behind the implemented algorithm is to calculate local displacements in the plane  $XY$  in the previous selected POIs as the maximum correlation between these POIs and a displacement in a neighbourhood. These displacements model the transformation locally because the movement of the tool does not significantly modify its orientation in this type of surgery. Finally, the algorithm resolves an optimization problem using these POIs and the displacements to calculate the global transformation.

In order to calculate the local displacements, the algorithm does as follows. Let  $P_j$  be one of these selected POIs with pixel coordinates  $(x_j, y_j)$ . First, the algorithm extracts a window  $\tilde{w}_{P_j}$  of size  $s_x \times s_y$  around the point  $P_j$  and a window  $\tilde{W}_{x_j, y_j}$  around the coordinates  $(x_j, y_j)$  in the image  $I_{i-1}$  of size  $S_x \times S_y$  with  $S_x > s_x$  and  $S_y > s_y$ . Then, these windows are scaled by a factor  $100/s_x \times 100/s_y$  to obtain  $w_{P_j}$  and  $W_{x_j, y_j}$ . This scaling factor is applied to obtain faster results in the local correlations. Then, the algorithm calculates the correlation of the windows  $w_{P_j}$  and all the subimages of size  $100 \times 100$  of the window  $W_{x_j, y_j}$ . Then, the displacement is computed as the coordinates of the maximum correlation. Finally, the mean  $m_{\Delta P}$  and standard deviation  $\sigma_{\Delta P}$  of all the displacements are calculated to erase the outliers by keeping all the POIs whose displacement verifies  $\Delta P_i \in \{m_{\Delta P} - 1.5\sigma_{\Delta P}, m_{\Delta P} + 1.5\sigma_{\Delta P}\}$ . In Fig. 5, the outliers are represented in yellow.

Once all the local displacements have been estimated, let  $(P_j, \Delta P_j)$  be the pair composed of the pixel coordinates and the displacement of point  $P_j$ . Then, the algorithm computes the best transformation  $\tilde{T}$  as the solution of the following minimization problem

$$\tilde{T} = \arg \min_T \sum_j \|T(P_j) - (P_j + \Delta P_j)\|^2$$

The non-linear least square problem is solved using the Lavenberg-Marquardt algorithm. Fig. 6 shows the final registration between two consecutively images.

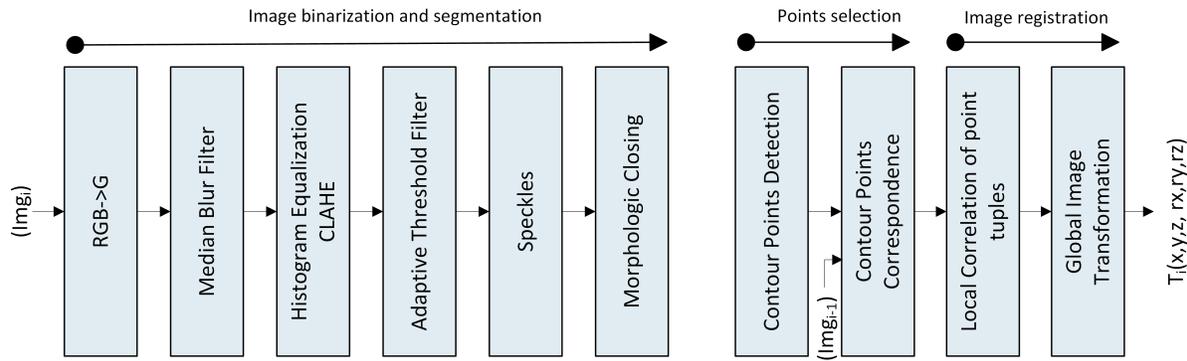


Fig. 1. Block schema of the tracking algorithm.  $Img_i$  and  $Img_{i-1}$  are the images at time  $t = i$  and  $t = i - 1$

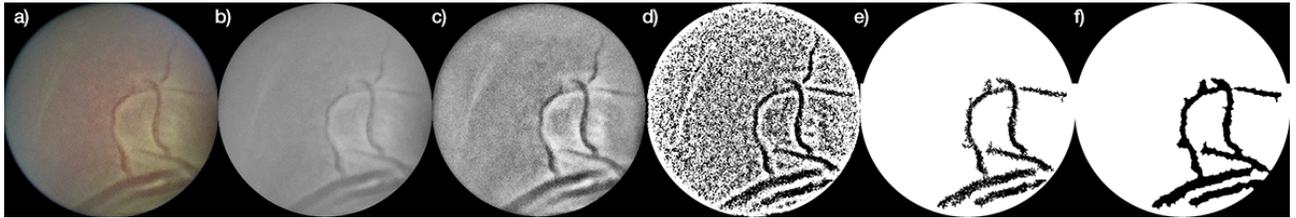


Fig. 2. Image binarization and segmentation sequence.

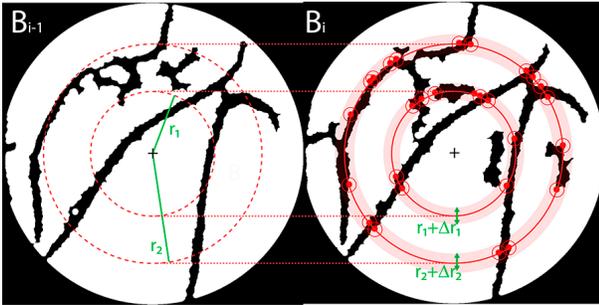


Fig. 3. POIs detection between the circles and the segmented area.

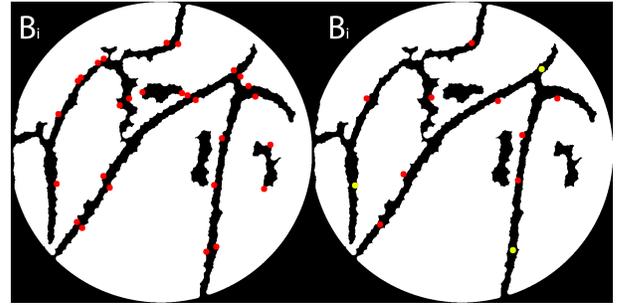


Fig. 5. Final selected POIs for image correspondences. The POIs that are marked as outliers after the local correlation step are represented in yellow.

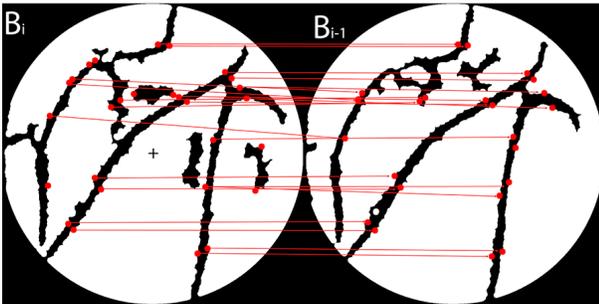


Fig. 4. POIs correspondences between  $B_i$  and  $B_{i-1}$ .

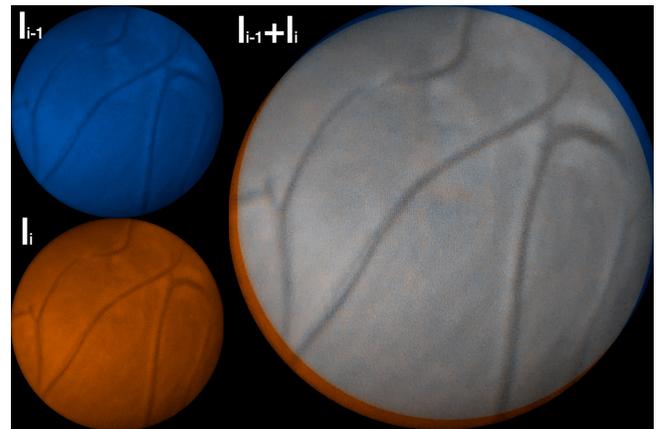


Fig. 6. Example of image registration of two consecutively frames.

## V. RESULTS

This sections presents the results of the described algorithm using a set sequences of real fetoscopic images, formed by sets of 20 to 40 frames. These frames have a size of  $963 \times 963$  RGB pixels with JPEG compression, which increases the image correlations difficulty. In order

to verify the performance of the algorithm (quality of image

registration), the similarity between the transformed  $I_{i-1}$  and the original  $I_i$  image has been calculated using a set of indexes: Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM). The MSE is defined as the mean square error per pixel. PSNR is defined as the maximum value of the reference image with respect to the MSE of the images. It is assumed that values over 20 dB indicates a satisfactory image pairing. SSIM calculates the structural degradation between two images.

The image sequences have been selected representing different parts of the placenta surface (different density of vascular structures, and background textures) in order to test the robustness and applicability of the algorithm. Fig. 8 shows the mosaics resulting from the different testing sequences. To generate these testing mosaics, the original image sequences have been used. To compute the transformed  $I_i$  image the result of the cumulative transformation  $AT_i = T_1 \cdots T_i$  is applied where  $T_j$  is the transformation between the image  $j - 1$  and  $j$ .

Table I presents the mean and standard deviation of these indexes for 5 different sequences. The algorithm shows a MSE of  $15.64 \pm 5.92$  using all the frames of the sequences. Likewise, calculating the PSNR using the same images, a value of  $28.82 \pm 1.30$  dB is obtained. The SSIM value for the same images is  $0.75 \pm 0.07$ , indicating a low structural information degradation.

Sequence 2 and 5 give worst results because sequence 2 is formed by a faster and random movements and, in sequence 5 some images show low vascular structures. In sequence 1 the camera describes a short and smooth trajectory, the results demonstrates the repeatability of the algorithm. Sequence 3 and 4 have high vascular structures and consequently the quality of the image registration is higher than in sequences 2 and 5, as expected. All the registrations suffer from the compression of the images which distortion the definitions of the edges and artificially varies the background texture of the placenta.

Finally, the error of the algorithm has been also evaluated as follows: first, the difference image between the previous frame and the transformation of the current frame. Afterwards, this image has been binarized using a threshold of 10. The mean of the error is 2.17% and a deviation error of 1.45%. Fig. 7 presents the percentage of error ones in the binarized difference image.

The complete frame computational time is bellow 25 ms in a serial configuration over CPU. The most time consuming step is the speckles filter. Initial tests in parallel configuration over GPU show an improvement of two times, which permit a real time image registration. To achieve optimized results a deeper study of parallel scalability is still necessary.

## VI. INTEGRATION INTO THE FETAL TELEOPERATION SYSTEM

This section presents the developed teleoperation station for fetal surgery and the integration of the vision based tracking system. Only the required details of the teleoperation

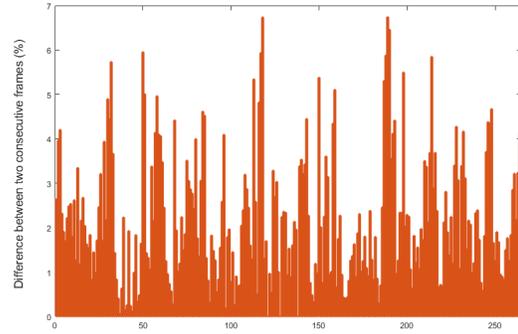


Fig. 7. Percentage of ones in a binarization of threshold 10 in the image difference between the previous frame and the current transformed image.

TABLE I  
MEAN AND STANDARD DEVIATION OF MSE, PSNR AND SSIM INDEXES OF 5 DIFFERENT SEQUENCES OF IMAGES.

| Test id     | Frames | MSE      |        | PSNR     |         | SSIM    |        |
|-------------|--------|----------|--------|----------|---------|---------|--------|
|             |        | mean     | std    | mean     | std     | mean    | std    |
| 1           | 27     | 10,1749  | 2,4666 | 30,397   | 0,9121  | 0,7041  | 0,0466 |
| 2           | 20     | 17,8535  | 8,5643 | 27,9842  | 1,6877  | 0,7767  | 0,0831 |
| 3           | 40     | 16,8433  | 5,2278 | 28,87    | 1,227   | 0,7359  | 0,0697 |
| 4           | 40     | 14,385   | 5,1529 | 28,8032  | 1,2801  | 0,7737  | 0,0607 |
| 5           | 29     | 18,9587  | 8,1879 | 28,0193  | 1,399   | 0,7588  | 0,0689 |
| All samples |        | 15,64308 | 5,9199 | 28,81474 | 1,30118 | 0,74984 | 0,0658 |

system are described, its complete description is out of the scope of this paper.

The fetoscopic teleoperation platform is based on a bilateral single master - single slave architecture. The master is composed of a 6DoF haptic device, multi-function pedals and an interactive user interface with several modules: fetoscopic view with augmented reality, an interactive map with information of the placenta and POIs (e.g. located anastomoses, umbilical insertions, equator, etc.) collected during the surgery, the teleoperation system status, etc. The slave is composed of a 6DoF robotic arm holding a 3mm fetoscope, an optical fiber extraction/retraction system for the coagulation laser and a trocar compensation system to actively control the depth of the trocar inside the uterus. Fig.9 shows a block schema of the teleoperation system.

Teleoperation system control is designed as a state graph and the states are defined by the FLP procedure: Idle, Free Navigation, Track Point, Trajectory to POI and emergency/safe fetoscope pose. Free navigation is designed to navigate over the placenta's surface to locate POIs. Trajectory automatically guides the fetoscope from current pose to a desired POI using free paths previously defined by the fetoscope. Track point actively maintains the center of the camera view, compensating the placenta's movements. This last state is used to study a potential POI, coagulate and review a POI. In all these three states, the vision system contributes to the fetoscope pose: the resulting slave pose is the sum of master device and placenta's movement (discriminating the robot

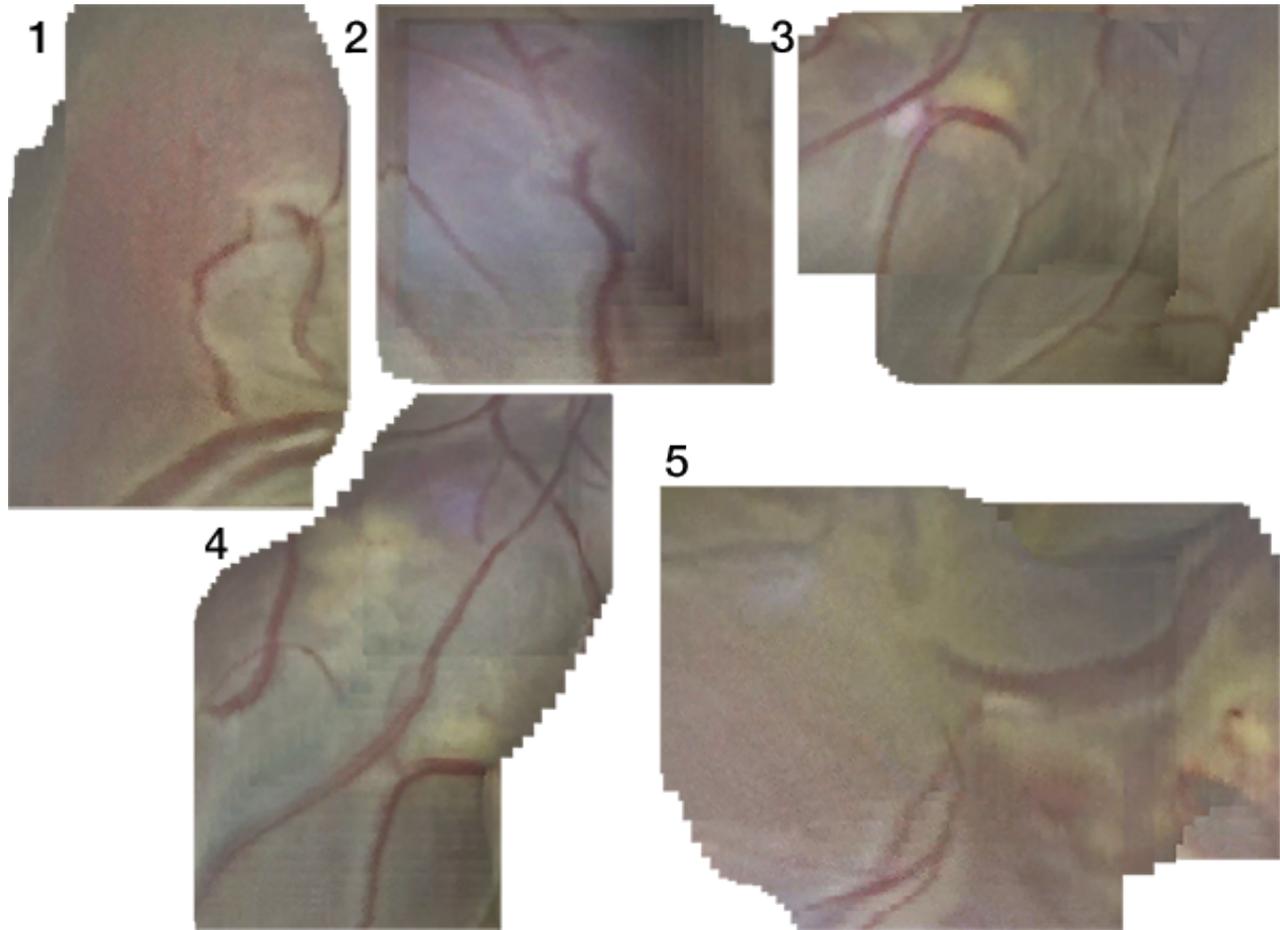


Fig. 8. Resulting placenta's surface reconstruction mosaics from the image sequences used in tests.

to placenta's variations due to master commands). Fig. 10 shows the state graph of the teleoperation system and all the possible transitions.

The vision system contributes to the fetoscope pose in all these three explained system status. In Free Navigation, the user can select a constant distance to placenta's surface for a safety navigation and to decrease the user's cognitive load. The vision system computes the fetoscope to placenta's surface distance and determines the fetoscope depth. In Trajectory status, the vision system keeps the minimum safety distance at each trajectory step. Once reached the destination POI, the vision system searches the matching between POI image and current image and actively guides the robot to keep the POI in the center of the fetoscopic view. Finally, in Track Point, the system recognises the current image and, again, actively guides the robot to fix the POI in the center of the view.

## VII. CONCLUSIONS

This paper presents a robust and accurate vision based tracking algorithm of the placenta's surface. The methodology allows dealing with the poor quality (noisy and blurred) of the images obtained with a real fetoscope and scenario. The method obtains the translation, camera axis rotation and

scaling from every two consecutive images containing visible vascular structures in real time.

The proposed algorithm can be used as image stabilization and fine localization of POIs during fetoscopic surgeries, like TTTS, over the placenta's surface. The inclusion of image stabilization increases the accuracy of anastomoses photo-coagulation, whereas the fine localization helps surgeons to perform a correct navigation, reducing the surgery time.

The next steps of this research will be the use of non compressed images extracted directly from the fetoscopic vision system and the tracking of fetoscope using a magnetic tracker (Aurora tracking system by NDI Medical). These tests will enable a direct comparison between fetoscope pose and algorithm results. The algorithm will be completely parallelized to reduce computational time. Finally, the complete teleoperation system will be tested in a real surgical scenario by the surgeons involved in the project.

## ACKNOWLEDGMENT

This work is part of the research of projects H2020-779813-SARAS (Smart Autonomous Robotic Surgeon) and "New high precision technologies in medicine and fetal surgery", funded by Cellex Foundation.

