



# Better Feature Matching for Placental Panorama Construction

Praneeth Satta<sup>1</sup>(✉), John A. Onofrey<sup>1,2</sup>, Mert O. Bahtiyar<sup>1,3</sup>,  
and Xenophon Papademetris<sup>1,2,4</sup>

<sup>1</sup> School of Medicine, Yale University, New Haven, CT 06520, USA  
{praneeth.satta, john.onofrey, mert.bahtiyar,  
xenophon.papademetris}@yale.edu

<sup>2</sup> Departments of Radiology and Biomedical Imaging,  
Yale University, New Haven, CT 06520, USA

<sup>3</sup> Obstetrics, Gynecology, and Reproductive Sciences,  
Yale University, New Haven, CT 06520, USA

<sup>4</sup> Biomedical Engineering, Yale University,  
New Haven, CT 06520, USA

**Abstract.** Twin-to-twin transfusion syndrome is a potentially fatal placental vascular disease of twin pregnancies. The only definitive treatment is surgical cauterization of problematic vascular formations with a fetal endoscope. This surgery is made difficult by the poor visibility conditions of the intrauterine environment and the limited field of view of the endoscope. There have been efforts to address the limited field of view of fetal endoscopes with algorithms that use visual correspondences between successive fetoscopic video frames to stitch those frames together into a composite map of the placental surface. The existing work, however, has been evaluated primarily on *ex vivo* images of placentas, which tend to have more visual features and fewer visual distractors than the *in vivo* images that would be encountered in actual surgical procedures. This work shows that guiding feature matching with deep learned segmentations of placental vessels and grid-based motion statistics can make feature-based registration tractable even in *in vivo* images that have few distinctive visual features.

**Keywords:** Feature matching · Fetoscopy  
Grid-based motion statistics · Mosaic construction  
Twin-to-twin transfusion syndrome

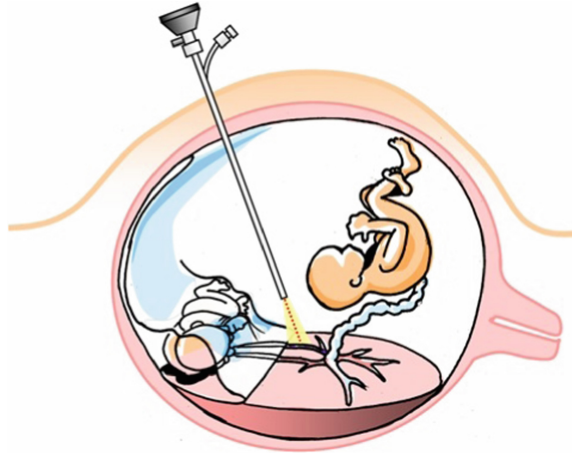
## 1 Introduction

### 1.1 Twin-to-Twin Transfusion Syndrome

Twin-to-twin transfusion syndrome (TTTS) is a disease of placental vasculature that can affect twin pregnancies. In some twin pregnancies, the two fetuses share a single placenta. It is possible for vascular connections to develop between the

portions of the placenta that serve each of the fetuses. When an unequal distribution of blood across these connections leads to a net flow of blood from one twin to the other, the result is TTTS [5]. TTTS can have serious consequences for both twins, including cardiac dysfunction in the twin that serves as a net blood recipient, injury to the central nervous system in the twin that serves as a net donor, and death in either twin [1, 5].

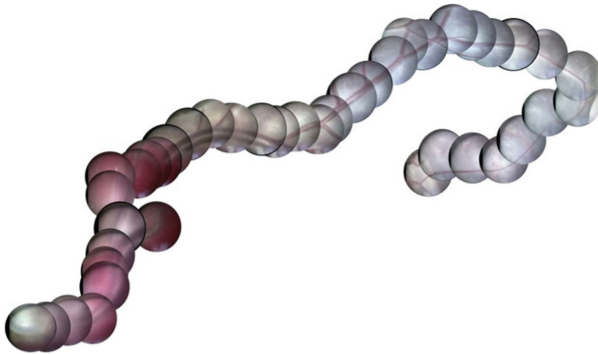
While there are several options for managing TTTS, there is only one definitive treatment: fetoscopic laser photocoagulation surgery [4]. In this procedure, a specialized endoscope known as a fetoscope is inserted through an incision in the maternal abdominal wall and then into the uterus. Once in the uterus, the fetoscope is used to inspect blood vessels on the surface of the placenta. Any problematic vascular connections that are found are cauterized with a laser. This procedure is illustrated in Fig. 1.



**Fig. 1.** A diagram of fetoscopic laser photocoagulation surgery for twin-to-twin transfusion syndrome by Luks [8]. Pictured are twin fetuses, each within their own amniotic sac. There is a single, shared placenta with problematic vascular connections that allow a net flow of blood from the donor fetus (left) to the recipient fetus (right). An endoscope (top) is used to inspect the placental vasculature and find problematic connections. When such connections are found, they are cauterized with a laser (center).

The challenges of fetoscopic laser photocoagulation are well described in the literature [12–14]. The problematic placental vascular formations cannot be visualized preoperatively with ultrasound or magnetic resonance imaging. They must therefore be identified intraoperatively using a fetoscope. This is made difficult, however, by the turbidity of amniotic fluid. The turbid nature of amniotic fluid not only reduces the clarity of the fetoscopic image, but also makes it impossible for the fetoscope’s attached light source to reliably illuminate structures that are more than a few centimeters away. The fetoscope must therefore be kept close to the placental surface, but this has the effect of reducing the field of view.

The distance across the placental vascular network (i.e. the distance from one twin’s umbilical cord to the other) can be several dozen times the diameter of the fetoscope’s field of view. As the surgeon can only see a small fraction of the placental surface at any given time, he or she must create a mental map of the relevant placental anatomy in real time and must rely on landmarks from this mental map in order to remain oriented as the surgery progresses. The high cognitive burden that fetoscopic laser photocoagulation surgery places on the surgeon increases the risk of error, which in the worst case can lead to the failure to identify and cauterize one or more vascular malformations, thereby necessitating a follow-up surgery. There has been interest in reducing the cognitive burden on the surgeon by replacing the surgeon’s mental map-making process with computer software that performs a similar task.

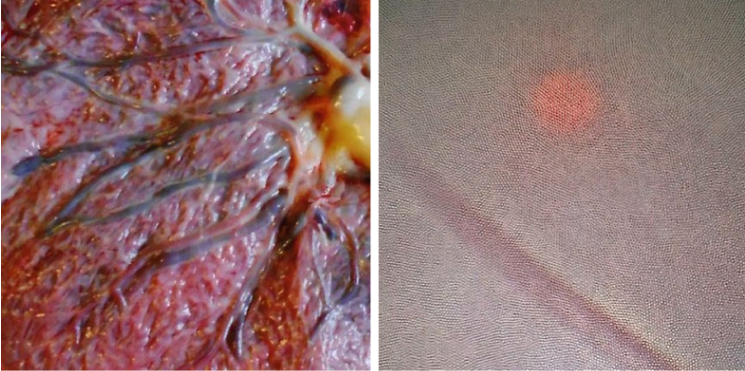


**Fig. 2.** An example of a panoramic view of the vasculature of a placenta that was created by concatenating fetoscopic video frames. This example was manually constructed from 30 min of fetoscopic footage from a fetoscopic laser photocoagulation surgery.

## 1.2 Prior Work

In the existing literature on placental panorama construction, by far the most common approach is to extract visual frame-to-frame correspondences and use those correspondences to calculate a homography from one frame to the other [3, 7, 10, 14]. Such approaches consist of a four step process: *(i)* using a feature detector to select key points from within an image; *(ii)* converting the high-dimensional raw pixel data of the image regions surrounding each key point into lower-dimensional vectors with the use of a feature description algorithm; *(iii)* matching the key points from one image with key points from the other, usually via a nearest-neighbor criterion on the key points’ associated feature descriptors; and *(iv)* calculating a homography from the coordinates of the matched key points. The two most popular feature matching and description algorithms in the existing literature on placental panorama construction are the Scale-Invariant Feature Transform (SIFT) and its derivation, Speeded Up Robust Features (SURF).

To the best of the authors' knowledge, all placental panorama construction studies to date have been evaluated primarily on *ex vivo* images [3, 10, 12–14] or images of placental phantoms [7]. *Ex vivo* images of placentas, however, tend to have more visual features and fewer visual distractors than *in vivo* images [6, 7]. Blood vessels are identifiable in both *ex vivo* and *in vivo* images, but *ex vivo* feature-rich backgrounds whereas *in vivo* images tend to have backgrounds that are almost entirely featureless (Fig. 3).



**Fig. 3.** Blood vessels are visible in both *ex vivo* and *in vivo* images of placentas. *Ex vivo* images, however, are rich in background features while *in vivo* images often have backgrounds that are entirely devoid of features. In the *in vivo* image, the guide light for the cauterization laser is visible in the upper center area. The guide light moves along with the fetoscope, so it is not suitable as a landmark for registration.

Gaisser *et al.* [7] simulated *ex vivo* and *in vivo* settings using a placental phantom and found that the performance of SIFT and SURF feature detectors could fall dramatically in the translation to *in vivo*. When applied to images from an *in vivo* setting with amniotic fluid of a yellow coloration, SIFT detected 73% fewer features than it did in an *ex vivo* setting. SURF detected 45% fewer features. The results reported by Gaisser *et al.* suggest that the underlying issue in registering *in vivo* placental images is a dearth of high-quality key points. If few key points are repeatable between different *in vivo* views of the same portion of a placenta, then there will be few matches. A homography calculated from a small number of matches will be highly sensitive to false or outlier matches. Furthermore, if the number of matches is low enough it will not be possible to compute a homography at all. Bian *et al.* [2] argue, however, that in many feature matching tasks, the underlying issue is not that there is a lack of good key points or good matches, but that standard matching techniques have difficulty distinguishing good matches from bad matches. It follows that better algorithms for determining matches between feature descriptors may be able to produce more accurate homographies for registering *in vivo* placental images into a panoramic map. In this work, we show that by extending the matching

algorithm beyond the typical nearest-neighbor approach, it is possible to extract meaningful matches between *in vivo* placental images even with low-quality key points and to exceed the accuracy of registrations produced with SURF and SIFT feature matching.

## 2 Methods

### 2.1 Feature Matching

Bian *et al.* [2] argue that when feature matching fails to produce sufficient matches, the underlying issue is often not a lack of good matches, but difficulty in distinguishing good matches from bad matches. In other words, when scoring matches (which is typically done by calculating the distance between the feature descriptors of the two matched key points), there tends to be a significant overlap between the score distribution of true matches and the score distribution of false matches. Setting a high minimum threshold for the match score minimizes the number of false positive matches but also eliminates many true matches.

Feature descriptor distance is not the only method for scoring matches. Bian *et al.* [2] propose scoring feature matches using the observation that true matches are likely to be neighbored by other true matches whereas false matches are more frequently found in isolation. Preliminary feature matches are first generated using the traditional nearest-neighbor approach. One image in the pair is then divided into a regularly spaced grid. A secondary score for a match that falls within the  $i$ -th cell of the first image and the  $j$ -th cell of the second is calculated as follows:

$$S_{i,j} = |X_{i,j}| - 1$$

where  $X_{i,j} = \{x_1, x_2, x_3, \dots, x_n\}$  is the union of matches found in the  $i$ -th cell of the first image and the  $j$ -th cell of the second. This secondary score is used to determine which cells in the first image are paired with which cells in the second. A constraint is then enforced in which key points within a given cell in the first image must match to its paired cell in the second image. Bian *et al.* refer to this approach as grid-based motion statistics (GMS). We apply a GMS match refinement step after the initial nearest-neighbor matching.

### 2.2 Feature Detection and Description

When matching key points with GMS, the quantity of key points is more important than their quality. We therefore use a feature detector that can generate a large number of key points: the AGAST corner detector [9]. We further increase the number of key points by lowering the AGAST detection threshold to zero and disabling the suppression of non-max corners. Although GMS is predicated on the notion that low quality key points can produce useful matches, it remains a fact that not all key points are of equal value. *In vivo* fetoscopic images are filled

with visual distractors such as the glare effects and floating debris in the amniotic fluid. These visual distractors are not useful for computing homographies between placental images.

In Sadda *et al.* [11], we showed that a neural network could be trained to segment blood vessels in *in vivo* placental images with human-level accuracy. We repurpose the segmentations produced by this trained neural network as a key point filter. Only key points that fall on a placental blood vessel are used; all other key points are discarded. The remaining key points are described with SIFT descriptors and matched with a nearest-neighbors approach. The matches are then refined with GMS.

### 2.3 Image Acquisition

*In vivo* placental images were acquired to evaluate the registration approach described in this paper. Intraoperative videos of ten fetoscopic laser coagulation surgeries performed at Yale-New Haven Hospital were obtained in a process approved by an institutional review board. All ten videos were recorded using a Karl Storz miniature 11540AA endoscope with incorporated fiber optic light transmission. 544,975 video frames were collected in total, accounting for approximately five hours of video. These video frames were cropped and downscaled from an initial resolution of  $1920 \times 1080$  pixels to a resolution of  $256 \times 256$  pixels.

## 3 Results and Discussion

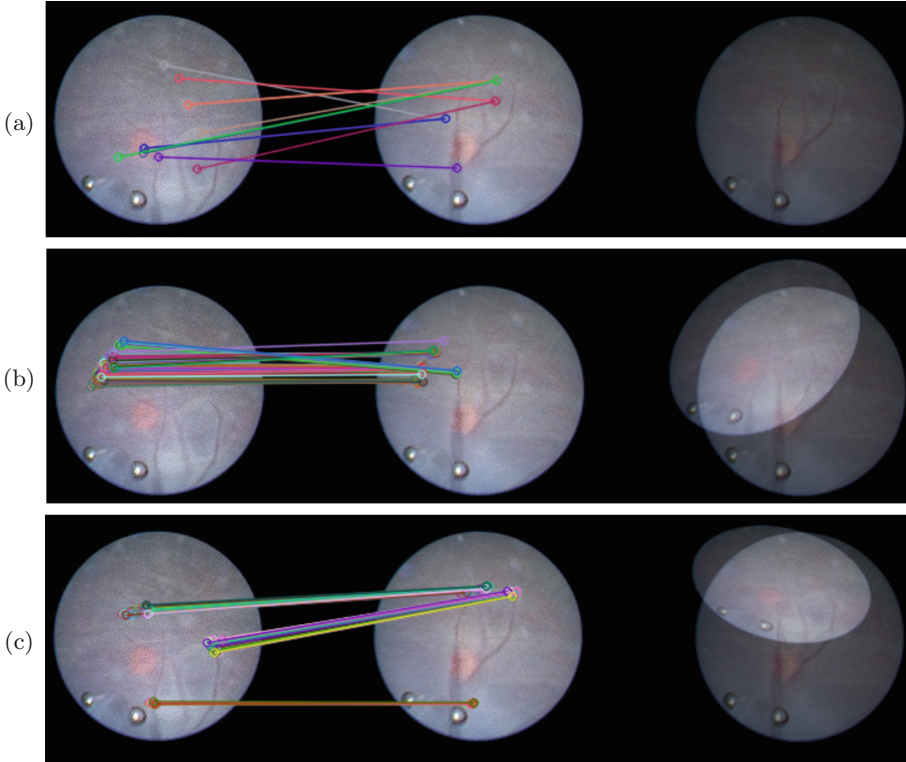
### 3.1 Synthetic Registration Task

188 video frames were extracted from the dataset of *in vivo* fetoscopic videos described in Sect. 2.3. Each image was randomly rotated between 0 and 360 degrees, translated by up to 64 pixels (one-quarter of the side-length of the viewport) along each axis, and perspective-warped by displacing each of the four corners of the image by up to 20 pixels.

**Table 1.** The results of the synthetic registration task described in Sect. 3.1. Fetoscopic video frames were distorted with randomly generated homographies. Various feature matching algorithms were used to recover the homographies. Each algorithm was evaluated in terms of success rate, defined as the percentage of image pairs for which the algorithm found enough matches to compute a homography, and transformation error, defined as the mean distance between a grid of points transformed by the ground truth homography and the same points transformed by the recovered homography.

Algorithm	Transformation error (pixels)	Success rate
SIFT	$143.3 \pm 366.2$	49.5%
SURF	$60.3 \pm 65.7$	85.1%
AGAST + SIFT + GMS	$3.2 \pm 5.5$	100.0%

Various feature matching algorithms were used to recover the homography between the original image and the distorted image. Each algorithm was evaluated in terms of success rate, defined as the percentage of image pairs for which



**Fig. 4.** An example from the natural registration task described in Sect. 3.2. The lower part of image  $A$  (left column) contains the upper ends of the blood vessels found in image  $B$  (center column). A composite image (right column) is created by overlaying the registered image  $A$  on top of image  $B$ . Several algorithms were compared: (a) Standard SURF key point detection and feature description yields a high ratio of false matches to total matches. This leads to image  $A$  being misregistered to such an extent that it falls completely outside of the composite image. (b) AGAST key point detection, SURF feature description, and GMS refinement of matches yields fewer false matches, but many of these matches are centered in a largely featureless background region. This leads to a number image  $A$  being registered to approximately the correct region of  $B$  but without proper alignment of the blood vessels in  $A$  to the corresponding vessels in  $B$ . (c) By using a deep-learned vessel segmentation algorithm, it is possible to limit AGAST key points to those that fall on blood vessels. This results in the algorithm correctly registering image  $A$  to the upper portion of image  $B$ . There are enough true matches for RANSAC-based homography estimation to identify and eliminate the false matches at the bottom of the images. The blood vessels in  $A$  are correctly aligned to the corresponding vessels in  $B$ .



the algorithm found enough matches to compute a homography, and transformation error, defined as the mean distance between a grid of points transformed by the ground truth homography and the same points transformed by the recovered homography. The results are summarized in Table 1.

The registration task in this experiment is admittedly trivial: since one image in each pair is a direct geometric transformation of the other image, a feature descriptor that lacked any invariance to lighting, illumination, or noise would in theory be able to generate matches across the images. However, this task is sufficient to show that the standard usage patterns of SIFT and SURF are unsuitable even for very trivial registration problems involving *in vivo* placental images. These methods fail to produce enough matches to compute a homography in a significant fraction of cases, and even when they can produce homographies, the homographies are of much lower quality than those produced by matching AGAST features with GMS.

### 3.2 Natural Registration Task

22 image pairs were selected from the dataset of *in vivo* fetoscopic videos described in Sect. 2.3. Each pair consisted of two images that depicted overlapping segments of the same vascular formation. To ensure that the frames were sufficiently different to make registration a nontrivial task, pairs were selected such that the video frames in each pair were acquired a minimum of 20 seconds apart. One image from each pair was manually rotated, translated, and perspective warped in an image editing program until it was aligned with the other image. The transformation

**Table 2.** The results of the natural registration task described in Sect. 3.2. Each algorithm was evaluated in terms of success rate, defined as the percentage of image pairs for which the algorithm found enough matches to compute a homography, and transformation error, defined as the mean distance between a grid of points transformed by the ground truth homography and the same points transformed by the recovered homography. The algorithms are as follows: (i) SIFT key point detection and SIFT feature description; (ii) SURF detection and description; (iii) SURF with key points filtered by a deep learned mask; (iv) SURF with a deep learned mask and with the Hessian threshold for detection reduced to zero; (v) AGAST feature detection, SIFT feature description and subsequent refinement of matches with grid-based motion statistics (GMS); and (vi) the AGAST/SIFT/GMS pipeline with the addition of a deep mask.

Algorithm	Transf error (pixels)	Success rate
SIFT	$97.1 \pm 34.6$	72.72%
SURF	$158.9 \pm 143.9$	100.00%
SURF + Deep Filter	$223.5 \pm 215.7$	40.90%
SURF (0 threshold) + Deep Filter	$118.9 \pm 57.0$	100.00%
AGAST + SIFT + GMS	$45.6 \pm 21.2$	100.00%
AGAST + Deep Filter + SIFT + GMS	$55.1 \pm 32.1$	100.00%



matrix corresponding to the concatenation of these editing operations was saved as the ground truth homography for that image pair.

Several feature matching and algorithms were executed on each image pair in an effort to recover the ground truth homography from visual correspondences. Each algorithm was evaluated in terms of success rate and transformation error, as defined in Sect. 3.1. The results are summarized in Table 2 and Fig. 4. Standard SIFT and SURF approaches perform poorly. SIFT fails to produce enough key point matches to produce a homography in over one quarter of cases. SURF is able to generate a homography more frequently, but the homographies that it produces have a high transformation error relative to the ground truth. One might expect that applying the deep learned vessel segmentations as a key point mask would help eliminate matches to visual distractors and increase match quality. However, applying deep filtering to SURF further reduces the number of available features, and lowering the Hessian threshold to increase the number of SURF features does not lead to better matches. Matching with GMS consistently produces the best registrations.

Adding a deep filter to GMS matching slightly increases the average transformation error. This is the result of images in which there is a single, linear blood vessel. As the deep filter limits key points to those that lie on a blood vessel, it causes the set of matched points in such images to be almost co-linear, and even slight deviations in the positions of matched key points can have a large effect on the computed homography if they are orthogonal to the axis of the lone blood vessel.

## 4 Conclusion

Prior research into the construction of panoramic maps of the placenta has made great strides in processing *ex vivo* placental images. Given that the ultimate goal is to use this technology intraoperatively, the next step is to extend existing techniques to handle the more complicated domain of *in vivo* images. However, the most common technique for panorama construction in the existing literature, nearest neighbor matching of SIFT and SURF features, gives unsatisfactory results even for very trivial registration tasks involving *in vivo* images. Feature matching with *in vivo* placental images is difficult because placental images lack a rich variety of visually distinct features. The appearance of one blood vessel on a placenta is not necessarily significantly different from the appearance of another blood vessel a centimeter away, and this leads to a high rate of false matches. In this work, we demonstrate that the paucity of visually distinct features is not necessarily a limiting factor in the registration of *in vivo* images. By using matching algorithms that impose a structure on matched elements – in this case a grid-based locality constraint – it is possible to significantly improve the quality of feature matches and the resulting image registrations.

**Acknowledgements.** This work was supported by the National Institutes of Health grant number T35DK104689 (NIDDK Medical Student Research Fellowship). The authors would like to thank Andreas Lauritzen for his assistance with data collection.

## References

1. Bahtiyar, M.O.: The North American fetal therapy network consensus statement: prenatal surveillance of uncomplicated monochorionic gestations. *Obstet. Gynecol.* **125**(1), 118–123 (2015)
2. Bian, J., Lin, W.Y., Matsushita, Y., Yeung, S.K., Nguyen, T.D., Cheng, M.M.: GMS: grid-based motion statistics for fast, ultra-robust feature correspondence. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2828–2837, July 2017
3. Daga, P., et al.: Real-time Mosaicing of Fetoscopic Videos Using SIFT, vol. 9786, pp. 9786–9786-7 (2016)
4. Emery, S.P., Bahtiyar, M.O., Moise, K.J.: The North American fetal therapy network consensus statement: management of complicated monochorionic gestations. *Obstet. Gynecol.* **126**(3), 575–584 (2015)
5. Faye-Petersen, O.M., Crombleholme, T.M.: Twin-to-twin transfusion syndrome. *NeoReviews* **9**(9), e380–e392 (2008)
6. Gaisser, F., Peeters, S.H.P., Lenseigne, B., Jonker, P.P., Oepkes, D.: Fetoscopic panorama reconstruction: moving from ex-vivo to in-vivo. In: Valdés Hernández, M., González-Castro, V. (eds.) *Medical Image Understanding and Analysis*, pp. 581–593 (2017)
7. Gaisser, F., Peeters, S.H.P., Lenseigne, B.A.J., Jonker, P.P., Oepkes, D.: Stable image registration for in-vivo fetoscopic panorama reconstruction. *J. Imaging* **4**(1), 24 (2018)
8. Luks, F.: Schematic Illustration of Endoscopic Fetal Surgery for Twin-to-Twin Transfusion Syndrome, December 2009
9. Mair, E., Hager, G.D., Burschka, D., Suppa, M., Hirzinger, G.: Adaptive and generic corner detection based on the accelerated segment test. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010. LNCS*, vol. 6312, pp. 183–196. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-15552-9\\_14](https://doi.org/10.1007/978-3-642-15552-9_14)
10. Peter, L., et al.: Retrieval and registration of long-range overlapping frames for scalable mosaicking of in vivo fetoscopy. *Int. J. Comput. Assist. Radiol. Surg.* **13**(5), 713–720 (2018)
11. Sadda, P., Onofrey, J., Imamoglu, M., Papademetris, X., Qarni, B., Bahtiyar, M.O.: Real-time computerized video enhancement for minimally invasive fetoscopic surgery. *Laparoscopic Endoscopic Robot. Surg.* **1**, 27–32 (2018)
12. Tella-Amo, M., et al.: A combined EM and visual tracking probabilistic model for robust mosaicking: application to fetoscopy. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016
13. Tella-Amo, M., et al.: Probabilistic visual and electromagnetic data fusion for robust drift-free sequential mosaicking: application to fetoscopy. *J. Med. Imaging* **5**(2), 5–16 (2018)
14. Yang, L., et al.: Towards scene adaptive image correspondence for placental vasculature mosaic in computer assisted fetoscopic procedures. *Int. J. Med. Robot. Comput. Assist. Surg.* **12**(3), 375–386 (2016)