Detector-Free Dense Feature Matching for Fetoscopic Mosaicking

Sophia Bano[1], Francisco Vasconcelos[1], Anna L. David[2], Jan Deprest[3], and Danail Stoyanov[1]

[1] *Wellcome/EPSRC Centre for Interventional and Surgical Sciences & Department of Computer Science, University College London, UK*
[2] *Fetal Medicine Unit, University College London Hospital, UK*
[3] *Department of Development and Regeneration, University Hospital Leuven, Belgium*
sophia.bano@ucl.ac.uk

INTRODUCTION

Twin-to-twin transfusion syndrome (TTTS) is a rare fetal anomaly that affects the twins sharing a monochronic placenta. It is caused by abnormal placental vascular anastomoses on the placenta, leading to uneven flow of blood between the two fetuses [1]. Fetoscopic Laser Photocoagulation (FLP) is used to treat TTTS, however, this procedure is hindered because of difficulty in visualizing the surgical environment due to limited surgical field-of-view, unusual placenta position, limited manoeuvrability of the fetoscope and poor visibility due to fluid turbidity and occlusions (Fig. 1. This adds to the surgeon's cognitive load and may result in increased procedural time and missed treatment, leading to persistent TTTS. Fetoscopic video mosaicking can create a virtual expanded field-of-view (FOV) image of the fetoscopic environment, which may support the surgeons in localizing the vascular anastomoses during the FLP procedure.

Classical video mosaicking techniques perform handcrafted feature detection, description (i.e. SIFT, SURF, ORB, etc) and feature matching in consecutive frames and homography estimation for image stitching. However, these methods perform poorly on the in vivo fetoscopic videos due to low resolution, poor visibility, floating particles and texture paucity or repetitive texture. Deep learning-based sequential mosaicking [2] method overcomes the limitation of feature-based mosaicking methods, but results in drifting error when stitching non-planar views. A recent intensity-based image registration [3] method relies on placental vessel segmentation maps for registration. This facilitated in overcoming some visibility challenges, however, this method fails when the predicted segmentation map is inaccurate or inconsistent across frames or in views with thin or no vessels. In the paper, we propose the use of transformer-based detector-free local feature matching (LoFTR) method [4] as a dense feature matching technique for creating reliable mosaics with minimal drifting error. Using the publicly available dataset [3], we experimentally show the robustness of the proposed method over the state-of-the-art vessel-based method.

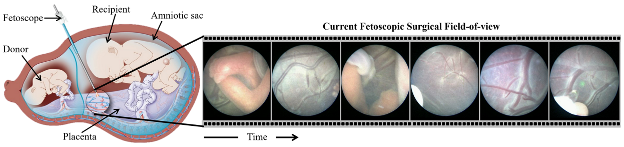

Fig. 1 During FLP, a fetoscope, having limited field-of-view, is inserted into the amniotic cavity and is used to localize and ablate the vascular anastomoses sites.

MATERIALS AND METHODS

A. *Detector-Free Feature Representation*

The recently proposed LoFTR [4] method first establish pixel-wise dense matches at a coarse level and later refine the good matches at a fine level. Given two consecutive frames F^t and F^{t+1}, a standard convolutional neural network architecture is used to extract dense features at coarse, $\tilde{F}^t, \tilde{F}^{t+1}$ (at $1/8^{th}$ of input resolution), and fine, $\hat{F}^t, \hat{F}^{t+1}$ (at $1/2^{th}$ of input resolution), levels from both frames. The coarse local features, $\tilde{F}^t, \tilde{F}^{t+1}$, becomes the input to the LoFTR module. LoFTR uses transformer with positional encoding, and self and cross-attention layers to transform $\tilde{F}_1, \tilde{F}_2$ into position and context dependent local features, denoted as $\tilde{F}^t_{tr}, \tilde{F}^{t+1}_{tr}$, that can be matched easily.

B. *Feature Matching*

The coarse level matches $\mathcal{M}_c$ between $\tilde{F}^t_{tr}, \tilde{F}^{t+1}_{tr}$ are established by using a differential matching layer, which gives a confidence matrix $\mathcal{P}_\rfloor$. The matches in $\mathcal{P}_\rfloor$ with confidence higher than a predefined threshold and that also satisfies the mutual nearest neighbour criteria are selected as $\mathcal{M}_c$. Finally, coarse ($\mathcal{M}_c$) to fine ($\mathcal{M}_f$) matches are obtained by taking local window size from fine-level features, $\hat{F}^t, \hat{F}^{t+1}$, at each coarse match positions, applying a LoFTR module to obtained the fine transformed representation and correlating them. For more detail, please refer to [4], in which it is shown that LoFTR produces high-quality matches even in regions having low-textures and are affected by motion blur or repetitive patterns; making it an ideal matching module for fetoscopic mosaicking.

C. *Registration and Mosaicking*

A circular mask covering only the fetoscopic FOV is used to remove matches from the unwanted blank re-

*This work was supported by the Wellcome/EPSRC Centre for Interventional & Surgical Sciences (203145Z/16/Z), EPSRC (EP/P027938/1, EP/R004080/1, NS/A000027/1, EP/P012841/1); and the Royal Academy of Engineering Chair in Emerging Technologies Scheme.

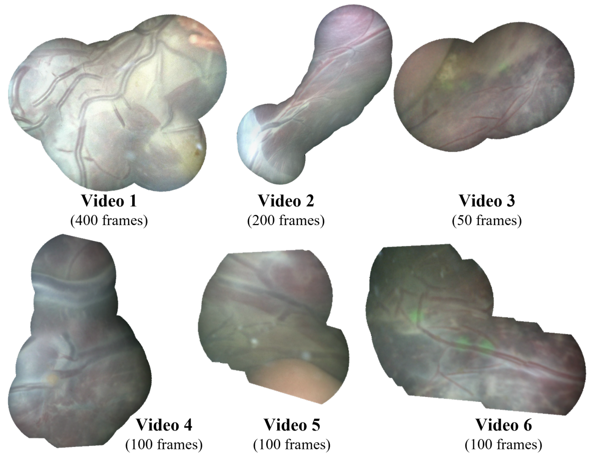

Fig. 2 Visualization of the generated mosaic using the proposed LoFTR-based method for the 6 in vivo clips.

gions. The registration between two consecutive frames, F^t and F^{t+1}, is approximated as an affine transformation [3] using RANdom SAmple Consensus (RANSAC) method. The obtained transformation is refined by using only the inliers with the Levenberg-Marquardt method that further reduces the reprojection error. Left-hand matrix multiplication is applied to the pairwise transformations to obtain the relative transformations of all frames in a video with respect to the first frame [2], following by image blending to generate an expanded FOV image.

RESULTS AND DISCUSSION

For the experimental analysis, we used the publicly available fetoscopy placenta dataset [3] that contains 6 in vivo fetoscopy video clips from 6 different TTTS procedures. The LoFTR matching model, pretrained on the ScanNet dataset [4], is used for obtaining the fine-level matches between two consecutive frames. Since groundtruth transformations are not available, we use the 5-frame structural similarity index measure (SSIM) presented in [3] for the quantitative evaluation of mosaics. The proposed LoFTR-based method is also compared with the state-of-the-art vessel-based [3] method (Fig. 3). The qualitative analysis on all 6 video clips is also performed (Fig. 2).

From Fig. 3, we can observe that the proposed LoFTR-based method performed significantly better on all video clips (except video 1) resulting in significantly low interquartile range and high median 5-frame SSIM when compared to the vessel-based [3] method. Video 1 contains heavy amniotic fluid particles dynamically floating in the in vivo environment, which affects the performs of the LoFTR resulting in inaccurate transformation estimation. In a vessel-based method, such particles are already filtered during vessel segmentation. Video 2-6 have dynamically changing non-planar views with some videos having low illumination and some frames having either no or very thin vessels. This negatively influences the vessel-based method, resulting in increase drifting error. LoFTR-based method, on the other hand, showed robustness even in regions having low-textures (very thin or no vessels) and low illumination. This is also evident from the qualitative results (Fig. 2 which can be compared with the qualitative results presented in [3]. In the case of Video 1, discontinuities are visible in the generated mosaic. In the case of Video 2-6, the generated mosaics are reliable and accurate without any visible discontinuities, which is also inline with the observations drawn from the quantitative results.

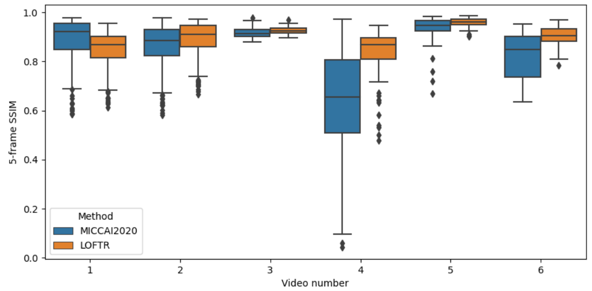

Fig. 3 Quantitative comparison of the proposed LoFTR-based method with the vessel-based (MICCAI2020) method [3] using the 5-frame SSIM metric.

CONCLUSIONS

We propose a fetoscopic video mosaicking method that benefited from the detector-free feature matching with transformers (LoFTR) [4] method, resulting in generating reliable virtual expanded field-of-view image of the intraoperative fetoscopic environment. Using the publicly available fetoscopy placenta dataset [3], we experimentally showed that the proposed LoFTR-based method outperformed the state-of-the-art vessel-based [3] fetoscopic mosaicking method. The proposed method is robust even in low-textured and low illumination non-planar views, which shows the potential of facilitating the surgeons during the TTTS procedure. Future work involves validating the proposed method on the larger FetReg [5] dataset.

REFERENCES

[1] A. Baschat, R. H. Chmait, J. Deprest, E. Gratacós, K. Hecher, E. Kontopoulos, R. Quintero, D. W. Skupski, D. V. Valsky, Y. Ville *et al.*, "Twin-to-twin transfusion syndrome (TTTS)," *Journal of Perinatal Medicine*, vol. 39, no. 2, pp. 107–112, 2011.

[2] S. Bano, F. Vasconcelos, M. Tella-Amo, G. Dwyer, C. Gruijthuijsen, E. Vander Poorten, T. Vercauteren, S. Ourselin, J. Deprest, and D. Stoyanov, "Deep learning-based fetoscopic mosaicking for field-of-view expansion," *International journal of computer assisted radiology and surgery*, vol. 15, pp. 1807–1816, 2020.

[3] S. Bano, F. Vasconcelos, L. M. Shepherd, E. Vander Poorten, T. Vercauteren, S. Ourselin, A. L. David, J. Deprest, and D. Stoyanov, "Deep placental vessel segmentation for fetoscopic mosaicking," in *Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 763–773.

[4] J. Sun, Z. Shen, Y. Wang, H. Bao, and X. Zhou, "LoFTR: Detector-free local feature matching with transformers," in *Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8922–8931.

[5] S. Bano, A. Casella, F. Vasconcelos, S. Moccia, G. Attilakos, R. Wimalasundera, A. L. David, D. Paladini, J. Deprest, E. De Momi, L. S. Mattos, and D. Stoyanov, "Fetreg: Placental vessel segmentation and registration in fetoscopy challenge dataset," *arXiv preprint arXiv:2106.05923*, 2021.