View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UCL Discovery

Real-Time Mosaicing of Fetoscopic Videos using SIFT

Pankaj Daga[a], François Chadebecq[a,b], Dzhoshkun I. Shakir [a], Luis Carlos Garcia-Peraza Herrera[a], Marcel Tella[a], George Dwyer[a,b], Anna L. David[c], Jan Deprest[d], Danail Stoyanov[b], Tom Vercauteren[a], Sebastien Ourselin[a]

[a]Translational Imaging Group, CMIC, University College London, UK,
[b]Surgical Robot Vision Group, CMIC, University College London, UK,
[c]Institute for Women's Health, University College London, UK,
[d]University Hospitals Leuven, Department of Obstetrics and Gynaecology, Leuven, Belgium

ABSTRACT

Fetoscopic laser photo-coagulation of the placental vascular anastomoses remains the most effective therapy for twin-to-twin transfusion syndrome (TTTS) in monochorionic twin pregnancies. However, to ensure the success of the intervention, complete photo-coagulation of all anastomoses is needed. This is made difficult by the limited field of view of the fetoscopic video guidance, which hinders the surgeon's ability to locate all the anastomoses. A potential solution to this problem is to expand the field of view of the placental surface by creating a mosaic from overlapping fetoscopic images. This mosaic can then be used for anastomoses localization and spatial orientation during surgery. However, this requires accurate and fast algorithms that can operate within the real-time constraints of fetal surgery. In this work, we present an image mosaicing framework that leverages the parallelism of modern GPUs and can process clinical fetoscopic images in real-time. Initial qualitative results on ex-vivo placental images indicate that the proposed framework can generate clinically useful mosaics from fetoscopic videos in real-time.

Keywords: Image stitching, Mosaicing, SIFT, image-guided surgery, GPU, CUDA, feature extraction

1. DESCRIPTION OF PURPOSE

Twin-to-twin transfusion syndrome (TTTS) is a complication that affects 9% of *monochorionic* identical twin pregnancies, where two or more fetuses share a common placenta.[1] The disease results from an imbalance in the blood circulation between the twins due to the presence of anastomoses in the monochorionic placenta. The implication of TTTS is that one of the twins (the *donor*) does not receive an adequate supply of blood for normal growth, while the other twin (the *recipient*) receives too much blood resulting in an overloaded cardiovascular system. Without intervention, the condition is often fatal for both twins. In advanced stages of TTTS, laser coagulation of the connecting vessels on the placenta between the twin fetuses can be a curative procedure. Under fetoscopic video guidance, the surgeon uses a laser fibre to photo-coagulate the blood vessels that connect the two fetuses. This involves visually inspecting the entire placenta to identify the crossing blood vessels that contribute to placental anastomoses. Laser energy is then used to coagulate these blood vessels, effectively separating the shared placenta and allowing each twin to develop independently.

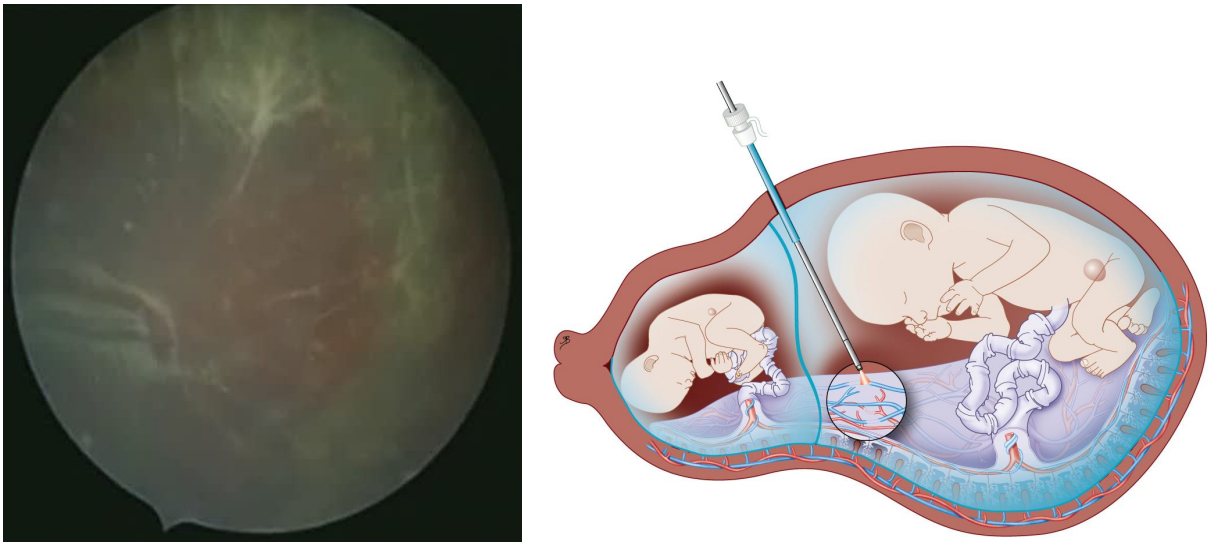
A key limitation of fetoscopic video guidance is the small field of view, which makes it difficult for the surgeon to visualize the full placental structure and ensure that all of the vascular anastomoses are identified and treated (figure 1). This problem may be overcome by creating a larger field of view through mosaicing of the fetoscopic video frames. The placental image mosaic may be helpful in locating the anastomoses and for spatial orientation during surgery.[2] These endoscopic image mosaics could also be fused with other imaging modalities to augment the information available during surgery.[3,4] The key to generating an accurate and usable image mosaic is to be able to extract accurate correspondences between video frames and stitch the video frames in real-time during clinical use.

This work proposes a graphics processing unit (GPU) based image mosaicing framework which can generate accurate image features and stitch fetoscopically acquired video frames in real-time. We demonstrate the potential of the proposed work on a synthetic dataset and highlight its clinical utility.

Send correspondence to Pankaj Daga. E-mail: p.daga@ucl.ac.uk

Medical Imaging 2016: Image-Guided Procedures, Robotic Interventions, and Modeling,
edited by Robert J. Webster, III, Ziv R. Yaniv, Proc. of SPIE Vol. 9786, 97861R
© 2016 SPIE · CCC code: 1605-7422/16/$18 · doi: 10.1117/12.2217172

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/06/2016 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx



a *b*

Figure 1. (a) A typical fetoscopic view during laser coagulation of the placental anastomoses. The field of view is small making spatial orientation and anastomoses localization challenging during the TTTS intervention. (b) A schematic representing the scenario during a TTTS procedure. The fetus on the right is the recipient and has received too much blood, while the fetus on the left is the donor and is deprived of an adequate blood supply. The fetoscope allows the viewing of a small region of the placenta and the surgeon has to visually identify all the anastomoses on the placental surface and coagulate them.

2. METHODS

SIFT[5,6] is one of the most widely used feature extraction and description algorithm in the computer vision community. The SIFT descriptor has the advantage of being invariant to translations, rotations and scaling transformations. It is also partially invariant to small perspective transformations and illumination changes. Although it is a widely used feature descriptor, its use in real-time applications is limited due to its computational complexity. In this section, we will describe the GPU implementation of the estimation of SIFT descriptor that enable its real-time use in a clinical setting. GPUs were traditionally used for the purpose of rendering images. However, they have revolutionized the field of parallel processing over the last years and are ubiquitous in the field of high performance computing. The GPU implementation in this work uses NVIDIA's CUDA API[7] and is the key component in the placental mosaicing pipeline.

2.1 SIFT Keypoint Detection

SIFT keypoints are detected using a cascade filtering approach to identify candidate keypoint locations. Detecting stable keypoints that are invariant to scale changes in the image involves looking for them across the entire scale-space. These keypoints are then described as local extremas in the *difference of Gaussian* (DoG) function convolved with the image. This involves smoothing and downsampling the image which can be done efficiently on the GPU[8] using fixed-sized separable kernels. To compute the local scale-space extrema in the DoG images, we make use of the *texture memory* on the GPU. The DoG images are mapped to the texture memory, which allows for extremely efficient lookup at each spatial location. Additionally, the GPU texture memory is optimized for data locality, making it a particularly efficient option to perform the extrema computation in a local window. This strategy is visualized in figure 2.

2.2 SIFT Orientation Computation

SIFT assigns one or more orientations to each of the detected keypoints based on local image gradient magnitude and directions. This step is computationally expensive as it involves representing image gradient information for all pixels in the neighbourhood of a keypoint using a histogram. A naive GPU implementation would be to use each thread in a block to process each keypoint. However, in the presence of a limited number of keypoints,

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/06/2016 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

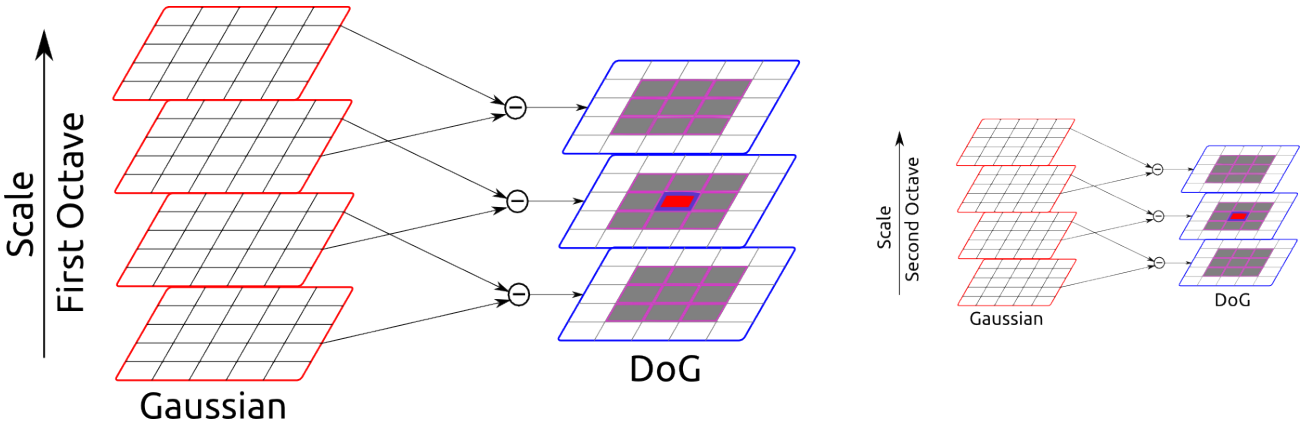


Figure 2. Figure shows the workflow for localisation of potential SIFT keypoints. The input image is convolved with Gaussian filters at various spatial scales and the difference of successive Gaussian filtered images are taken. In order to determine if a given pixel is an extrema, it is compared with its 26 nearest neighbours as highlighted in the image. This difference of Gaussian images are mapped to the GPU texture memory to take advantage of the fact that GPU texture caching mechanism is optimized for spatial locality. This allows for efficient lookup within the neighbourhood of a pixel leading to a computationally efficient way of detecting local extremas.

this leads to a sub-optimal GPU occupancy. We propose to use each thread in a block to process a single bin in the histogram related to a given SIFT keypoint. We make use of atomic operations, provided by CUDA, to increment the shared histogram between the threads of a block. This strategy to maximize the GPU occupancy leads to a speed-up of over 30 times over the naive GPU implementation. This strategy is visualized in figure 3.

2.3 SIFT Descriptor Computation

The final step is the computation of the SIFT descriptor which characterizes each of the keypoints as a 3-D spatial histogram of the image gradients. This descriptor aims to achieve robustness by encoding the image information in a localized set of gradient orientation histograms. The gradient is sampled around the keypoint location using a Gaussian weighting to give less weight to the gradients that are far away from the keypoint center. Once all histogram entries are computed, they are concatenated to form a 128-dimensional descriptor vector. An approach, similar to the computation to the keypoint orientations is used to compute these descriptors on the GPU where we aim to maximize the GPU occupancy by having each thread in the block process one pixel around the keypoint neighbourhood. However, the size of the keypoint neighbourhood can easily exceed the maximum number of threads in a CUDA thread block. To alleviate this problem we break the computation down into multiple chunks, which are spread across multiple passes within the execution of the block.

2.4 Estimation of Spatial Transformation and Visualization

SIFT descriptors are computed for every frame and corresponding features between adjacent frames are identified through brute force distance matching between the descriptor vectors. Robust homography estimation is obtained by using a GPU implementation of the RANSAC[9] algorithm. The estimated transformation is used to update the global mosaic through fast image resampling on the GPU. Efficient blending and updating of the mosaic is obtained by sharing graphics memory buffer between the display and compute parts of the algorithm avoiding unnecessary data copy.

3. RESULTS

3.1 Computational Time

The experiments were conducted on a computer with 16 CPU cores, 32 GB of RAM and running a 64-bit Linux system running Ubuntu 14.04. The GPU accelerated code was executed on an NVIDIA Quadro K4000 graphics card with 3GB of graphics memory. The computational time comparison is made against VLFeat,[10] a widely used software suite for computer vision algorithms. In table (1) we compare the time taken for SIFT feature

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/06/2016 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

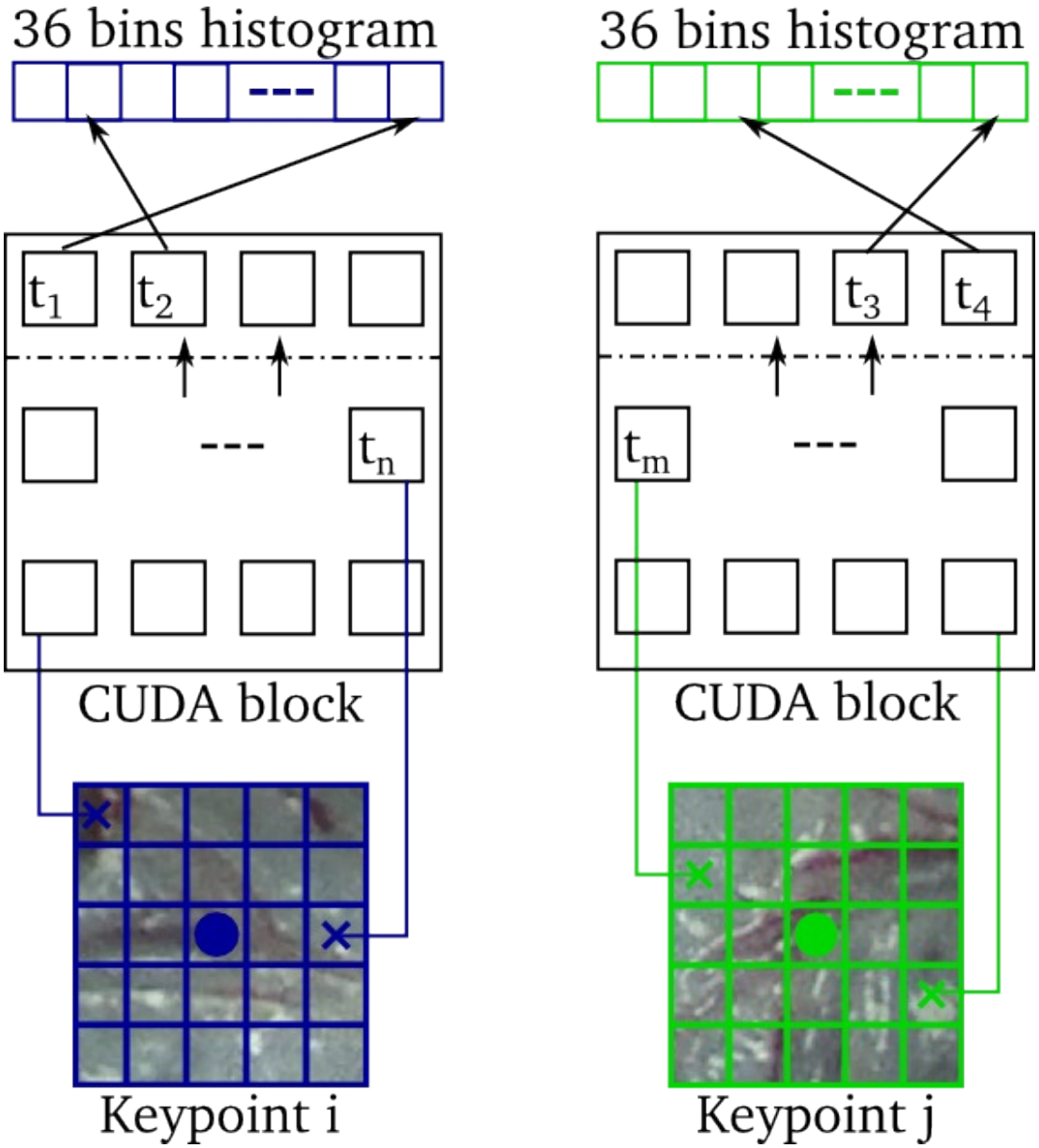


Figure 3. Figure shows the strategy used to maximize the GPU occupancy during keypoint orientation computation. Each thread in a given GPU/CUDA thread block operates on a single pixel in the neighbourhood of a keypoint. GPU atomic operations are used to synchronize the updates to a shared 36 bin histogram between the threads of a block. Once, the histogram is computed, the first 36 threads of a block are used to smooth and find the maximas in the histogram. Finally, a single thread in the block is used to compute up to 2 orientations for a keypoint. This results in a speed-up of about 30 times over the naive implementation where each GPU thread processes a single keypoint and thus results in sub-optimal GPU occupancy for a typical fetoscopic image.

detection and descriptor computation for images of various sizes. All times are reported in milliseconds. Even in the case of high resolution images (1920×1080), the proposed framework achieves SIFT computation at 10 frames/sec. The reader should note that the clinical images are usually cropped to include the region of interest and these images are typically much smaller size as shown by the highlighted row in table (1) and can be processed in real-time on most modern GPUs using the proposed method. Additional steps which compute the optimal transformation between successive frames and update the mosaic image are also implemented on

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/06/2016 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

the GPU. While the computation time of resampling the image using trilinear intepolation and updating the global mosaic image is negligible, the computation time involving the transformation estimation is related to the number of SIFT keypoints detected between the frames. During our experiments, the mean additional time taken to compute the spatial transformation and update the mosaic was approximately 14 ms for clinical images.

Image Size	VLFeat	Proposed
1920 x 1080	4318(321)	95.43(0.9)
1024 x 768	2125(217)	40.97(1.6)
800 x 600	887(112)	30.98(0.74)
640 x 480	643(98)	20.43(0.62)

Table 1. Mean(standard deviation) time (in milliseconds) taken to compute the SIFT features using VLFeat and the proposed framework when using images of various sizes. The use of GPU enables computation of SIFT features in real-time for clinical images. The processing workflow typically involves cropping the image to only include areas of interest and decrease computation time. A typical clinical video frame size after cropping is shown by the highlighted row.

3.2 Qualitative Assessment on Phantom Data

We also performed a preliminary quality assessment of the algorithm on phantom data acquired using the setup described in the following section. The initial results look promising as seen by examining the computed image mosaic in figure 4. The spiral motion of the robotic arm was reconstructed with good fidelity.

3.2.1 Data Acquisition Setup

The phantom data acquisition setup is highlighted in 4(a). The phantom data was acquired by mounting the shaft of the fetoscope to the flange of the seven DoF robot arm (Kuka LBR iiwa 7 R800) using a custom made component. The arm was then programmed to follow an expanding spiral trajectory approximately parallel to the image plane, with a maximum radius of 60 mm. The phantom consisted of an image of a term human placenta which was collected after a caesarean section delivery following normal pregnancy.

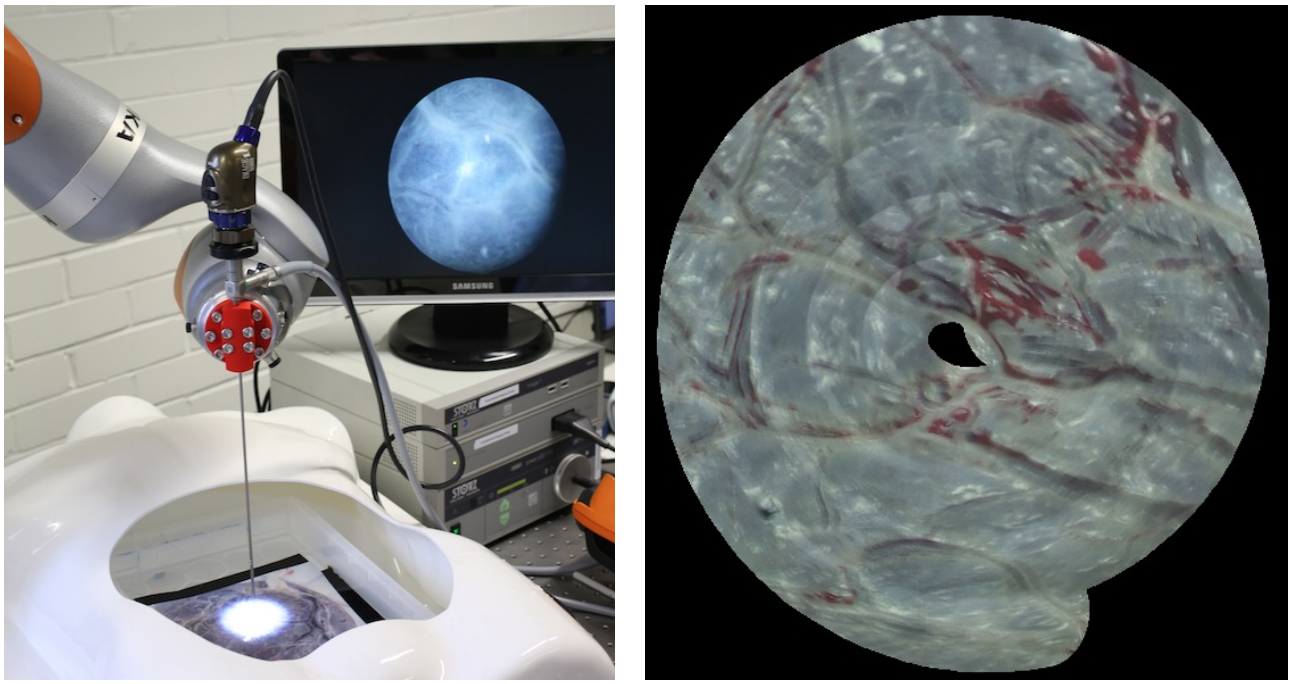
4. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

The paper presents a GPU implementation of a real-time image mosaicing framework which is geared towards minimally invasive fetal surgery. Initial results show that the framework can operate well within the time constraints of fetal surgery and could potentially generate surgically useful placenta mosaics in real-time during the intervention. This has the potential to be applied to the clinic and generate accurate image mosaics while the surgeon is performing the surgery. We believe this could lead to more successful interventions by minimizing the chances that vascular anastomoses are left untreated during the intervention. The proposed framework is developed under strict software quality control with the aim of translation to surgical use through our clinical collaborators. We aim to release the algorithms under a open source licence in the near future.

5. CONCLUSIONS AND FUTURE WORK

The paper highlights the implementation details of the fetal surgery mosaicing framework. While the initial qualitative results look promising, future work will aim to do quantitative validation using both synthetic and real clinical datasets. We also aim to build upon this framework to explore other research questions with regards to image mosaicing during fetal surgery. This includes development of more realistic transformation models, robust estimation of feature correspondences and even new feature descriptors that may be more specialized for use in fetal surgery. The proposed framework's modular design will enable one to change and add various components with minimal software coding effort. Additionally, this work has the potential to lead into future research where these anastomoses are detected automatically from the generated placental mosaic. We believe this could reduce surgery time, result in improved instrument navigation during surgery and complement the surgeon in the task of identification and accurate localization of all the surgical targets.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/06/2016 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx



a *b*

Figure 4. (a) The imaging setup for the synthetic experiment. The fetoscope is mounted on a Kuka robotic arm (Kuka AG, Germany) using a custom joint. The arm was programmed to traverse a spiral trajectory while recording the video of the placenta phantom. (b) The placental image mosaic. The mosaic is computed by stitching together overlapping small field of view images. The initial results show good coherence in the vascular structure visible in the stitched image.

6. ACKNOWLEDGEMENTS

This work was supported through an Innovative Engineering for Health award by Wellcome Trust [WT101957]; Engineering and Physical Sciences Research Council (EPSRC) [NS/A000027/1]. Jan Deprest is being funded by the Fonds voor Wetenschappelijk Onderzoek Vlaanderen(FWO; JD as clinical researcher 1.8.012.07). Anna L. David is supported at UCL/UCLH by funding from the Department of Health NIHR Biomedical Research Centres funding scheme. Danail Stoyanov receives funding from the EPSRC (EP/N013220/1, EP/N022750/1), the EU-FP7 project CASCADE(FP7-ICT-2913-601021) and the EU-Horizon2020 project EndoVESPA (H2020-ICT-2015-688592). Sebastien Ourselin receives funding from EPSRC (EP/H046410/1, EP/J020990/1, EP/K005278) and the MRC (MR/J01107X/1). Marcel Tella, George Dwyer and Luis Herrera are supported by the EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging (EP/L016478/1). We would also like to thank NVidia Corporation for the donation of the GeForce Titan card used in this research.

REFERENCES

[1] Lewi, L., Jani, J., Blickstein, I., Huber, A., Gucciardo, L., Mieghem, T. V., Doné, E., Boes, A.-S., Hecher, K., Gratacós, E., Lewi, P., and Deprest, J., "The outcome of monochorionic diamniotic twin gestations in the era of invasive fetal therapy: a prospective cohort study," *American Journal of Obstetrics and Gynecology* (2008).

[2] Reeff, M. and Székely, G., "Mosaicing of endoscopic placenta images," in [*Informatik 2006. Informatik für menschen*], (2006).

[3] Liao, H., Tsuzuki, M., Mochizuki, T., Kobayashi, E., Chiba, T., and Sakuma, I., "Fast image mapping of endoscopic image mosaics with three-dimensional ultrasound image for intrauterine fetal surgery," *Minimally Invasive Therapy and Allied Technologies* (2009).

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/06/2016 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx

[4] Yang, L., Wang, J., Kobayashi, E., Liao, H., Sakuma, I., Yamashita, H., and Chiba, T., "Ultrasound image-guided mapping of endoscopic views on a 3d placenta model: A tracker-less approach," in [*Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions*], (2013).
[5] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision* (2004).
[6] Brown, M. and Lowe, D. G., "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vision* (2007).
[7] Nickolls, J., Buck, I., Garland, M., and Skadron, K., "Scalable parallel programming with CUDA," *Queue* (2008).
[8] Podlozhnyuk, V., "Image convolution with CUDA," *NVIDIA Corporation white paper, June* (2007).
[9] Fischler, M. A. and Bolles, R. C., "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM* (1981).
[10] Vedaldi, A. and Fulkerson, B., "VLFeat: An open and portable library of computer vision algorithms," (2008).

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/06/2016 Terms of Use: http://spiedigitallibrary.org/ss/TermsOfUse.aspx