Laparoscopic Bipolar Forceps Positioning in Stereo Video for Robotic Surgery Training

Akihiro Terashima
Osaka University
akihiro-terashima@hiel.ist.osaka-u.ac.jp

Ando Hide
Osaka University
hide@ist.osaka-u.ac.jp

Michael Wu
University of California Berkeley
253 Cory Hall, Berkeley, CA 94720
wuxiaohua1011@berkeley.edu

Abstract

In this paper, we will explore methods for tracking exact object position from a stereo video source, which involves estimation of 3 dimensional position and using methods that relies heavily on the open-sourced library OpenCV rather than the SLAM algorithm. Specifically, this method relies heavily on StereoSGBM class for depth estimation and the CSRT Tracker for object tracking(Although there are effort in using neural network such as MRCNN for tracking). The proposed method addresses two basic problems with no prior knowledge about the camera specification – 2 dimensional position tracking, and detecting depth information from unknown camera source.

1. INTRODUCTION

Laparoscopy is a low-risk, minimally invasive procedure that requires only small incisions. However, there is a shortage of experienced doctor and low-cost training systems. Dr. Naohiro and Prof. Ando et. al [11] have proposed a re-experience training system using the recorded video of the actual surgery and superimpose the trainees forcep onto the video – to ensure that the trainees forcep can move just as the experienced doctors.

Dr Naohiro and Prof. Andos [11] research had a significant impact on training the new doctors placement of forcep in 2 dimensions, however, the system could be improved if depth is also added for a complete 3 dimensional simulation of the surgery.

2. Related Work

Yihong Wu's paper [12] have a great summary on the current progress and trends in image based localization. As they have stated, image based camera localization is the key to tasks such as virtual reality, augmented reality, robotics, autonomous driving. However, utilizing this technology in the medical field, specifically in to the laparoscopic surgery.

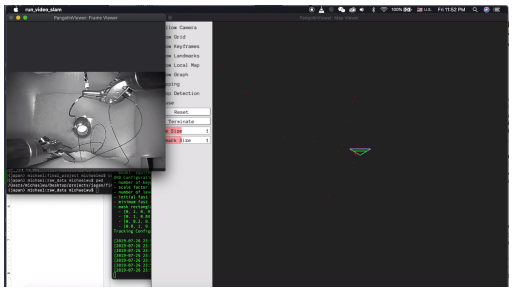

Figure 1. OpenVSLAM result with Our input video as source

Much work in camera position estimation under anonymous environments have been done with SLAM, or Simultaneous Localization and Mapping, which can help construct unkonw environments from videos in real time and online. Furthermore, there are different methods of SLAM, monocular slam, multicular SLAM, Multi-kind sensors SLAM, and Learning SLAM. This paper will not survey nor compare different SLAM algorithms, however, this paper is heavily related to multi-cular SLAM, since we are receiving an input video from a pair of stereo camera.

Some influential stereo SLAM proposed includes S-PTAM [8], which can compute the real scale of the map and overcome the limitation of PTAM for robot navigation. ORB-SLAM2 [7] is a system for both monocular, stereo, and RGBD cameras, it includes map reuse, loop closing and relocalization capabilities.

However, after consideration, we decided that SLAM algorithms requires substantial camera movements to de-

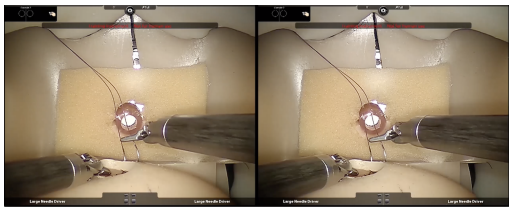

Figure 2. Raw input Stereo Video Demo

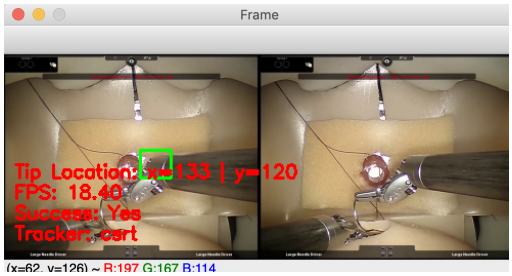

Figure 3. Original OpenCV CSRT Tracker Demo

termine camera position. In our case, we are unable to afford such movements. And therefore, we must turn to purely image based 3D reconstruction. [Figure **??**] Work by D.Scharstein et. al [9] describes a method for acquiring high accuracy depth map using structured light that does not require calibration of the light sources. Pascal Fua describes a parallel stereo algorithm for dense depth map calculations [2] H. Mohan Chand used OpenCV and Stereo Camera to create point cloud. [4]

There are other more complex medical systems. Fuchs et. al [3] presented the design and prototype of a three-dimensional visualization system to assist with laparoscopic surgical procedures. The system also uses epth extraction from laparoscopic images. They noted that much work remains for it to be clinically useful, notably in speed, reconstruction and registration of 3D imagery. Danail Stoyanov et. al [10] proposed a strategy for dense 3D depth recovery and temporal motion tracking for deformable survace, especially for laparoscopic surgery.

3. Method

All the source code and data are open sourced and published. The link can be found on `https://github.com/wuxiaohua1011/stereo_image_to_point_cloud_japan`

OpenCV is used for efficient depth map generation. Our input is a stereo video stream. [Figure: 2] Our over-arching agenda is to track object's position first in 2D , then detect its corners, and then find the depth information for its corner. For tracking object's 2D position, we have attempted both OpenCV's image tracker CSRT and using neural network such as MRCNN [1]. Current implementation is using OpenCV's CSRT tracker [Figure: 3], which is based on Alan et. al's Deiscriminative Correlation Filter with Channel and Spatial Reliability [6]

For every frame, we cropped the stereo video into left and right images, both images of the same dimension. For our input source, our parameters are as Table 1

However, to facilitate visualization, we will only overlay the results onto the left image, specifically the results includes tracking box and the depth information.

For extracting depth image, we depend heavily on OpenCV's StereoSGBM class, which is based on

Left Image	min	max	difference
X	60	1260	1200
Y	60	910	850

Right Image	min	max	difference
X	1340	2540	1200
Y	160	910	850

Table 1. Image Boundary Statistics

Hirschmuller's algorithm [5] with slight modification. The difference between StereoBM algorithm and StereoSGBM should be noted – StereoBM perform Winner takes it all(WTA) and sum of absolute difference(SAD) between left and right images, and thus compute the minimum local cost for each pixel. While StereoSGBM, in addition to computing the least local optimal disparity value, also enforces smoothness constraints between neighboring pixels, which tends to lead to better results. Answer adapted from OpenCV forum

StereoSGBM will compute the disparity image and return it as a NumPy array. Thus for every frame, we have a distinct disparity image. However it should be noted that different ways of using StereoSGBM will generate different results. In the source code, two different functions

```
# option 1
disparity = calculate_disparity_michael(imgL,
imgR, show_disparity=True)

# option 2
disparity = calculate_disparity_tim(imgL,
imgR, show_disparity=True)
```

Option 1 [Figure 4] uses the same StereoSGBM objects for both left and right images, however, the disparity map it generate is less smooth than Option 2. However, for option 2 [Figure 5, 6], despite its smoothness, it is unusable because it generates a depth map of different scale for each frame, possibly due to its dependency on time according to the author Timotheos Samartzidis This paper is going to continue discussion with option 1.

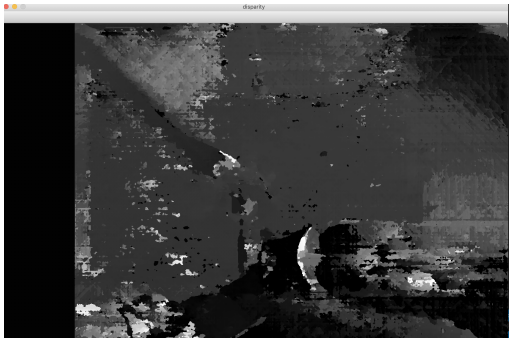

Figure 4. Option 1 Disparity Map

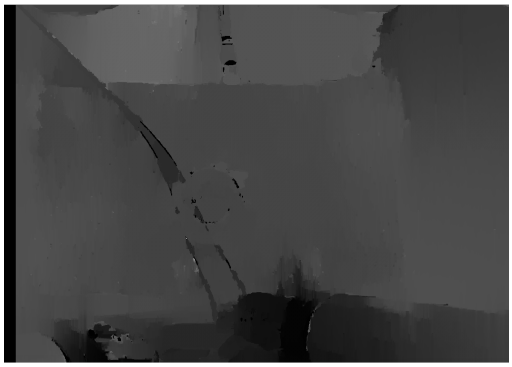

Figure 5. Option 2 Disparity Map Frame 1

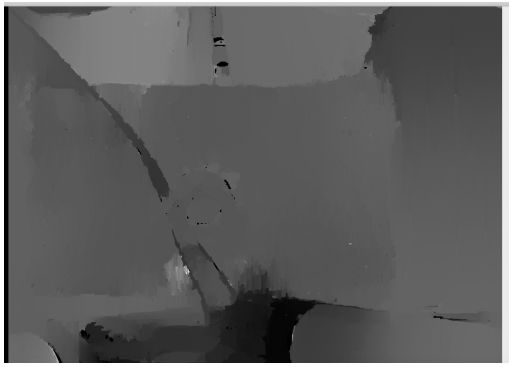

Figure 6. Option 2 Disparity Map Frame 2

After generating the disparity map, we estimated the camera parameter and used the triangulation equation

$$depth = (baseline * focal)/disparity$$

. We made a slight modification, since we are going to be working with a selected area's depth, we are going to compute the average of the depth using the average of the disparity value in that area. Or in simple words, compute the average of the disparity value in the bounding box that the tracking algorithm mentioned above generates, put it through the triangulation equation.

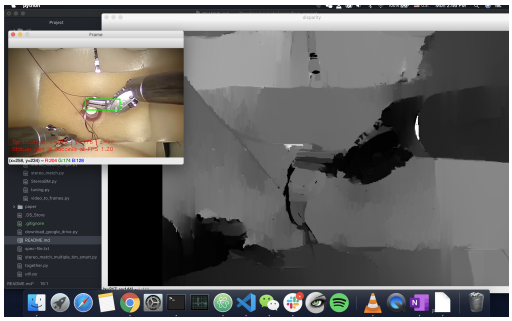

Figure 7. Sample Program run

After all information are generated, we display the depth map, bounding box, video frames, and other information. [Figure 7]

4. Conclusion

This paper is an initial result in using OpenCV for precise laparoscopic forcep position detection. More information needs to be known, such as the exact specification of the camera. And after getting these specifications, thorough tests needs to be run to see if our depth map is indeed a one-to-one relationship with the actual distance, In simple words, if d is the depth output, is there a variable K or one-to-one function $f(d)$ such that $K * d$ or $f(d)$ will yield the real world coordinate given the current output.

Another possible upgrade would be to transfer the current code into C++, and use GPU or OpenMP, or other methods for run-time optimization. Currently, the code has been tested and ran on MacBook Pro with 2GHz Intel Core i5 and 8GB of memory.

References

[1] P. Burlina. Mrcnn: A stateful fast r-cnn. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 3518–3523, Dec 2016.

[2] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, 6(1):35–49, Dec 1993.

[3] H. Fuchs, M. A. Livingston, R. Raskar, D. Colucci, K. Keller, A. State, J. R. Crawford, P. Rademacher, S. H. Drake, and A. A. Meyer. Augmented reality visualization for laparoscopic surgery. pages 934–943, 10 1998.

[4] M. C. Hanumara. Calibrating and creating point cloud from a stereo camera setup using opencv. 08 2016.

[5] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341, Feb 2008.

[6] A. Lukezic, T. Vojir, L. Cehovin Zajc, J. Matas, and M. Kristan. Discriminative correlation filter with channel and spatial reliability. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[7] R. Mur-Artal and J. D. Tardós. ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. *CoRR*, abs/1610.06475, 2016.

[8] T. Pire, T. Fischer, J. Civera, P. De Cristforis, and J. J. Berlles. Stereo parallel tracking and mapping for robot localization. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1373–1378, Sep. 2015.

[9] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 1, pages I–I, June 2003.

[10] D. Stoyanov, A. Darzi, and G. Z. Yang. Dense 3d depth recovery for soft tissue deformation during robotically assisted laparoscopic surgery. In C. Barillot, D. R. Haynor, and P. Hellier, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004*, pages 41–48, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[11] A. Terashima, A. Morishima, K. Obama, Y. Sakai, T. Maeda, and H. Ando. Laparoscopic sigmoidectomy surgery training system using AR follow-up experience of real human surgery. In *International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, ICAT-EGVE 2018, Posters and Demos, Limassol, Cyprus, November 7-9, 2018*, pages 19–20, 2018.

[12] Y. Wu. Image based camera localization: an overview. *CoRR*, abs/1610.03660, 2016.