Improving Stability of Vision-based Camera Tracking by Smartphone Sensors

Jaejun Lee*, Kei Obata*, Maki Sugimoto*†, Hideo Saito*†
* Graduate School of Science and Technology, Keio University Yokohama, Japan
† CREST, Japan Science and Technology Agency

ABSTRACT

3D tracking is a trending issue in the field of augmented reality, which brings several challenges in a variety of situations, such as estimating the camera poses in dim conditions, obstructed scenes. It is difficult to stabilize the pose estimation results, especially when the result is dependent on only camera images under occlusions. However by using inertial sensors in smartphones, we can obtain image-independent pose estimation. In this study, we propose a simple sensor fusion method which supports image-based 3D camera pose estimation method. We demonstrate tracking results that is more robust in an obstructed scene compared to the image-based tracking method

Keywords: pose estimation, 3D object tracking, sensor fusion, generative learning.

1 INTRODUCTION

Recent years, 3D tracking has become a significantly popular issue in computer vision. 3D tracking [1] allows the system to estimate the position and orientation of a target object or camera from the input images. Many of the significant methods of 3D tracking considers the usage of natural feature points.

Keypoint matching is one of the key methods of 3D tracking which extracts local feature and makes descriptors from the input image. There are several algorithms for this task: Scale-Invariant Feature Transform (SIFT) [2] and Speed-Up Robust Features (SURF) [3] are popular methods. These methods has the invariance to rotation and translation, however, they are still vulnerable to large changes on perspective transformation. This causes failure when matching the keypoints between the input images as some keypoints have less robustness to large changes on perspective transformation. To maintain robustness to large perspective changes, View generative learning (VGL) [4] offers improvement in the robustness of viewpoint changes by selecting stable keypoints from various viewpoints. However, in certain cases, camera pose estimation fails due to the lack of stable keypoints by occlusions in actual situations. In this study, we propose an improved 3D camera pose estimation method by combining inertial measurement unit (IMU) information of a smartphone and the image based camera tracking method.

2 IMPROVING STABILITY OF IMAGE-BASED CAMERA TRACKING BY SMARTPHONE SENSORS

In image based tracking methods, it is possible to estimate the camera pose by finding natural keypoints that matches in camera images. However, when the visible keypoints lacks due to the occlusion or dynamic lighting condition, there are case that the tracking fails.

On the contrary to image tracking, orientation information from the IMU sensor is independent from the input image. By considering the sensor information with the difference between the images, the estimation result of image based tracking improved under hard conditions. Figure 1 shows a snapshot of a tracking result of our camera pose estimation method.

* jaejunlee@imlab.ics.keio.ac.jp

2.1 Camera Tracking by VGL

In this study, we improve stability of an image based tracking algorithm: viewpoint generative learning [5] by combining sensor information. Generative learning is divided into two parts: learning and tracking. In learning phase, generative learning extracts keypoints from images synthesized at various viewpoints for textured 3D model, which should be somehow obtained beforehand, such as using 3D modeling tool(e.g. 123D catch in this experiment). Using the 3D model, we can collect keypoints on the 3D model that are repeatedly detected in various viewpoints, which we call them as stable keypoints. Each stable keypoint usually have different SIFT descriptors in different viewpoint images because of viewing directions are different. Therefore, we record all SIFT descriptors for each stable keypoint in database. In tracking phase, we compare each detected keypoint from input image and stable keypoints that are stored in the database. We compare the descriptors and find the nearest descriptor by fast approximate nearest neighbor (FLANN) [6] matching. After comparing both keypoints, we match those values and find correspondences of them with robust estimator RANSAC. In our method, descriptors are sorted in order of the identification number of stable keypoints. Using these sorted descriptors finally estimates the camera poses.

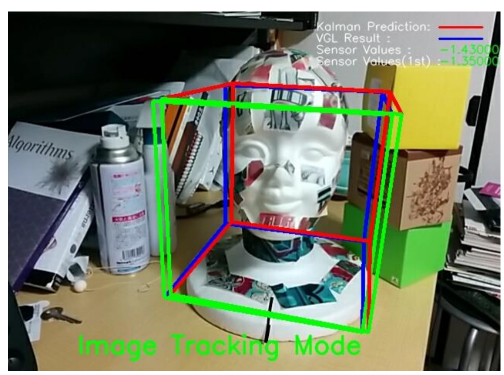


Figure 1: A snapshot of our camera pose estimation method based on View generative learning and IMU information.

2.2 Applying IMU information

The unique feature of our method is its capability to estimate the camera pose by the combination of the image based tracking method and IMUs such as Accelerometers and Gyroscopes on a smartphone. Since most of smartphones have monocular cameras and IMUs, our improved method can be applied for various scenes.

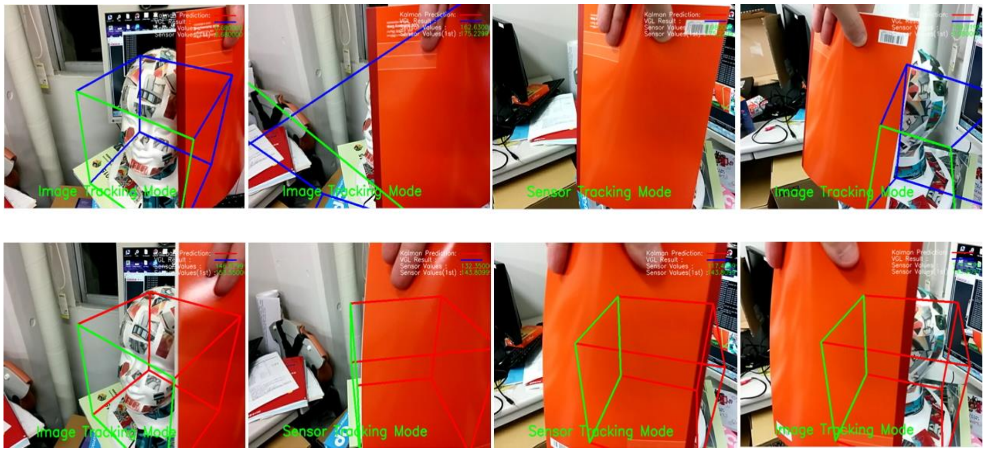
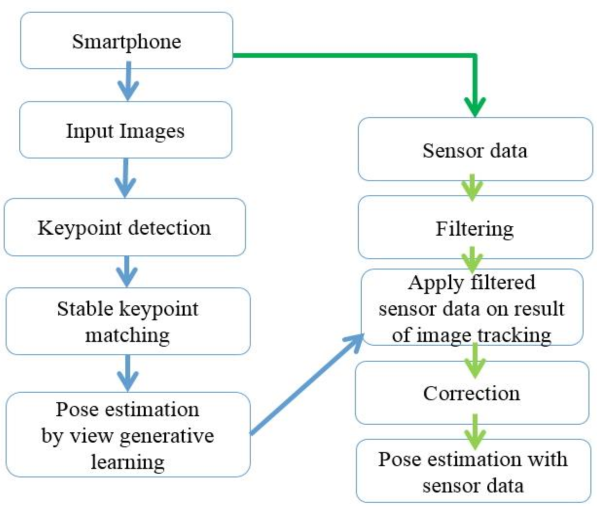


Figure 2: 3D tracking results with the camera pose estimation under occlusion (Top : without IMU, Bottom : with IMU)

To apply IMU information to the image based tracking, it is important to identify the accurate results of the image based tracking. We classify the tracking results by considering the standard deviation of the camera pose changes of the tracking for short time. When a stable result is observed, our method simply relies on the result. However, when the tracking result is classified as inaccurate, our method switches to sensor-based tracking. In this phase, IMU information is integrated with the latest reliable image-based tracking result. Figure 3 depicts the flow of our proposed method.

Accelerometer and gyroscope information is filtered and integrated with a Kalman Filter. This integrated sensor information shows the rotation and translation of the tracking device. When image-based tracking is not available, we estimate the current camera translation and position in a relative coordinate from the latest reliable result using the sensor information. After obtaining a new reliable tracking result, our method updates the relative coordinate system according to the tracking result.



3 Demonstration

In the demo, we show the improved stability of the image-based tracking algorithm and an application which shows the possibility of our method. The application estimates the camera pose in real time through the improved image based tracking method. Our application is able to estimate camera poses and overlay Computer Graphics objects. When an obstacle appears in front of camera, the image-based tracking algorithm cannot estimates the camera pose. In such a case, our improved tracking method is able to show a stable camera pose estimation result. The image and the IMU sensor data taken from the smartphone is sent to the host computer. In the host computer, extraction of keypoints and matching between the images are carried out and the tracking result is calculated. The tracking result is then sent back to the smartphone. Figure 2 shows the tracking results under occlusion.

In the demonstration, we use the following devices: Host Computer (Windows 7 Professional 64 bits, Intel Core i7-2700K 3.50GHz CPU, 8.00GB RAM and NVIDIA GeForce 550Ti GPU) and Smartphone (Google Nexus 5). Also, we use SiftGPU to extract keypoints from camera images.

4 Conclusion

In this study, we proposed a method to improve stability of an image-based camera tracking method by combining IMU sensor information of a smartphone. The method is able to show a stable tracking performance on occluded keypoints. We demonstrate an application to show possibility of our method.

References

[1] Park, Y., Lepetit, V., Woontack, W.: Multiple 3D Object tracking for augmented reality. In: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 117.120 (2008)

[2] D. G. Lowe.: Distinctive image features from scale-invariant keypoints. Int. J. Computer Vision 60(2), 91.110 (2004)

[3] Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-Up Robust Features (SURF). Comput. Vis. Image Underst. 110(3), 346.359 (2008)

[4] Yoshida, T., Saito, H., Shimizu, M., Taguchi, A.: Stable keypoint recognition using viewpoint generative learning. In: Proceedings of the 8th International Conference on Computer Vision Theory and Applications, vol. 2, pp. 310–315, 2013

[5] Yukiko Shinozuka, Francois de Sorbier and Hideo SaitoSpecular 3D Object Tracking by View Generative Learning Proc. 2014 Irish Machine Vision and Image Processing (IMVIP2014), pp.9-14, (2014)

[6] M. Muja., D. G. Lowe.: "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration", in International Conference on Computer Vision Theory and Applications (VISAPP'09), (2009)