Reduced egomotion estimation drift using omnidirectional views

Yalin Bastanlar

Computer Eng. Dept., Izmir Institute of Technology, Izmir, Turkey
yalinbastanlar@iyte.edu.tr

Estimation of camera motion from a given image sequence becomes degraded as the length of the sequence increases. In this letter, this phenomenon is demonstrated and an approach to increase the estimation accuracy is proposed. The proposed method uses an omnidirectional camera in addition to the perspective one and takes advantage of its enlarged view by exploiting the correspondences between the omnidirectional and perspective images. Simulated and real image experiments show that the proposed approach improves the estimation accuracy.

Introduction: Estimating the 3D motion of a camera (egomotion) and the 3D scene structure simultaneously from a sequence of images is known as simultaneous localization and mapping (SLAM) in robotics community and as structure from motion (SfM) in computer vision community. This problem has been studied extensively over the years and different approaches were proposed. However, satisfactory results are not easily obtained due to the high dimension of the space of unknowns and challenges regarding the nature of the work such as erroneous feature matching or degenerate cases.

Many methods to perform the described task start by estimating the geometry of two views. The 3D structure built with these two views is used to estimate the pose of a new view with the help of the matches between the new view and already reconstructed points. More cameras are added in this manner. The reconstructions are often improved using an iterative optimization, so called bundle adjustment. Due to their sequential nature, these methods suffer from drift (error built up) in estimation of camera poses (egomotion). The drift removal approach presented in [1] is based on the assumption that image sequence forms a loop. Another loop closing method is given in [2]. However, real life applications do not necessarily contain loop sequences. Therefore, accurate egomotion estimation is crucial for the sequential algorithms.

In this letter, we show that adding omnidirectional views to a SfM consisting of only perspective views improves the accuracy of egomotion estimation. With its 360° view, an omnidirectional camera can provide us corresponding points with virtually all perspective views. Our hypothesis is that these point correspondences help to obtain more accurate camera motion estimation. We investigate, with experiments, the effect of adding an omnidirectional view to a perspective-only SfM by measuring the drift in both cases.

One can think of employing only an omnidirectional camera to fully exploit the large field of view advantage. However, due to low resolution and distortion, matching of feature points is less reliable than perspective cameras. In an omnidirectional visual odometry study [3] (the term ‘odometry’ indicates the work is concentrated on the position estimation of the camera), authors point out that the rotation estimated from features alone gives rise to large errors. Solely for rotation estimation they convert the omnidirectional image to panoramic image and estimate the shift in the image to extract the angle of rotation. It was also mentioned that when the feature points are distributed in an unbalanced fashion, estimation is degraded. In our study, we use the omnidirectional camera as an aid to the perspective camera approach.

Our method: Since our approach involves using omnidirectional cameras with perspective ones, we employ a mixed camera SfM framework [4]. Our method employs the sphere camera model [5], which is able to cover single viewpoint omnidirectional cameras as well as perspective cameras. Following an automated hybrid feature point matching and estimation of epipolar geometry, 3D points are reconstructed with iterative linear triangulation. After the structure and motion estimation with the first two frames, new frames are added one by one using the matches between the new view and already reconstructed points. Finally bundle adjustment (BA) is performed using Levenberg-Marquardt method, minimization criterion of which is the 2D reprojection error.

Experiments with simulated images: We first analyze the proposed approach in a simulated environment. Generated points are randomly distributed in a 3D space. Simulated perspective camera sees only a part of the scene points at a certain position. As the camera moves, some of the previously viewed points are lost and new points appear in the image. The omnidirectional camera views all the points and it is positioned at the half of the distance traveled by the perspective camera (Fig. 1). Gaussian noise (with σ=1 pixel) is added to pixel coordinates.

We measured the egomotion estimation drift for changing number of views (frames) to compare the three cases; (i) initial estimation without bundle adjustment (BA), (ii) after perspective BA and (iii) after mixed camera BA (Fig. 2). Please note that the same number of BA iterations was applied in both perspective and mixed cases and the experiments were repeated 50 times and the average drift was taken.

Fig. 2a shows the drift (in cm.) for the compared approaches when the traveled distance by the perspective camera is fixed to 3 meters. We observe that changing the number of views within a fixed distance does not affect the initial drift. Although one may expect the drift to increase with more views, the estimation of motion becomes more successful with closer views due to the increased number of matched points. More importantly, bundle adjustment with the omnidirectional view (mixed BA) outperforms the perspective camera approach. The performance difference decreases with increasing number of views.

Fig. 2b shows the measured drifts when the traveled distance by the perspective camera increases 60 cm with each view. This time, the drift increases with the addition of new views. Mixed BA consistently gives better results than perspective-only BA, which indicates that adding an omnidirectional camera improves the accuracy of egomotion estimation.

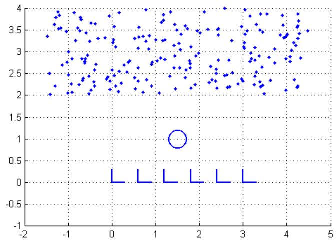

Fig. 1 *A top-view of the simulated environment. Perspective camera is moving to the right. Omnidirectional camera is indicated with a circle.*

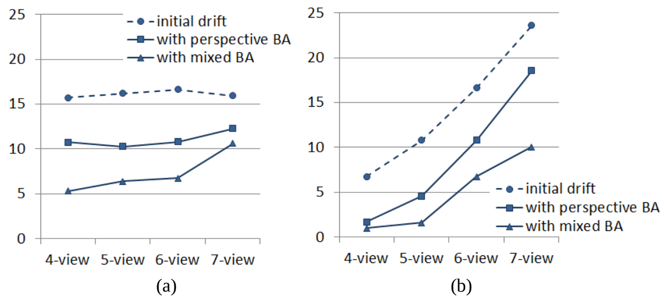

Fig. 2 *Egomotion estimation drift (y axis) for compared approaches and changing number of views (x axis). (a) The distance traveled by the perspective camera is fixed. (b) The distance increases with each view.*

Experiments with real images: We performed an experiment where the positions of the camera were recorded while capturing the images of the scene, which enabled us to compare the estimated camera positions with the real ones. One omnidirectional and six perspective images used in this experiment are shown in Fig. 3. Scene does not change much as the perspective camera moves but since the viewing angle varies, baseline increases and matched points change as in the case of simulated images.

Fig. 4a shows the initial estimation of the structure and camera motion. The camera movement is indicated with arrows and numbers. The actual distances between all six consecutive views are equal to 40 cm, however a drift can easily be observed in the estimated motion towards the end of the sequence. Fig. 4b shows the result after perspective-only BA and Fig. 4c shows the result when the omnidirectional camera is added (mixed BA).

Fig. 3 *Omni and perspective images of the first real image experiment.*

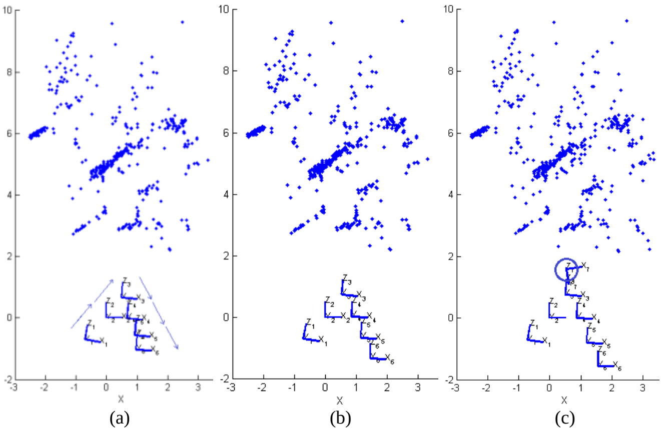

Fig. 4 *Structure and motion estimated from images in Fig.3. Figures show top-view and the z axes of the estimated perspective camera views are toward the reconstructed scene points. (a) The initial estimation with perspective-only SfM. The results after bundle adjustment for (b) perspective-only BA and (c) mixed BA. The omnidirectional camera is indicated with a circle around it, z axis of which is looking down.*

Since the reconstructions are correct up to a scale factor, we normalized the measured drift using the known real world distances. Initial estimation drift is 26.7 cm whereas the drifts after perspective BA and mixed BA are 14.2 cm and 6.5 cm respectively. Result depicts a performance increase by the addition of the omnidirectional view.

An easy way to measure the drift is using a sequence of images that form a closed loop since the drift is the distance between the estimated positions of the first and last views. We performed such an experiment with six views. Fig. 5a shows the initial estimation, the order of the views and the drift which corresponds to 20.55 cm after scale normalization. Fig. 5b shows the perspective-only BA result (drift=18.05 cm) and Fig. 5c shows the result with the omnidirectional view (drift=0.95 cm). The improvement gained by mixed BA is higher than the average performance difference found in simulated image experiments. Note that, we pretended not to know that there is a closed loop and did not perform point matching between the first and the last frames because this normally requires a loop detection algorithm such as [2]. Here, we used the loop sequence to measure the drift easily.

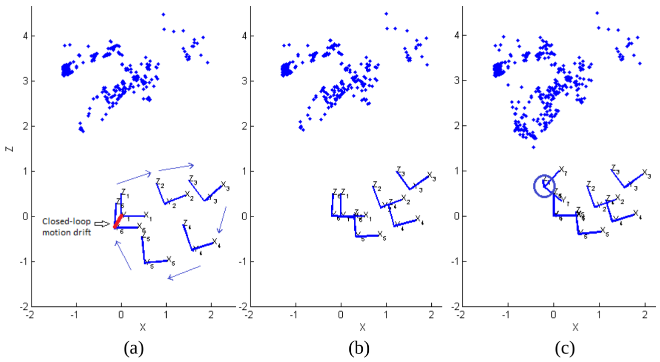

Fig. 5 *SfM experiment with a closed loop sequence. (a) The initial estimation, the drift is depicted in red. (b) The result after perspective-only BA. (c) The result after mixed BA. The drift is almost disappeared.*

Conclusion: We demonstrated that adding an omnidirectional camera to the perspective camera approach of egomotion estimation decreases the drift, which is an important problem especially for sequential algorithms. In this letter, we did not work on egomotion estimation over long distances, rather we concentrated on SLAM and SfM which definitely requires an accurate camera pose estimation. However, our method can also be valuable for visual odometry over long distances [6]. Mobile applications are also possible since omnidirectional viewing apparatus became available at low-cost even for smart phones.

References

1 Cornelis, K., Verbiest, F., and Van Gool, L.: ‘Drift detection and removal for sequential structure from motion algorithms’. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(10), pp. 1249-1259.

2 Ho, K.L., and Newman, P.: ‘Detecting Loop Closure with Scene Sequences’, *Int. Journal of Computer Vision*, 2007, **74**(3), pp. 261-286.

3 Scaramuzza, D. and Siegwart, R.: ‘Appearance-Guided Monocular Omnidirectional Visual Odometry for Outdoor Ground Vehicles’, *IEEE Transactions on Robotics*, 2008, **24**(5), pp. 1015-1026.

4 Bastanlar, Y., Temizel, A., Yardimci, Y., and Sturm, P.: ‘Multi-view Structure-from-Motion for Hybrid Camera Scenarios’, *Image and Vision Computing*, 2012, **30**(8), pp. 557-572.

5 Geyer, C. and Daniilidis, K.: ‘A unifying theory for central panoramic systems’, European Conference on Computer Vision (ECCV), 2000.

6 Nister, D., Naroditsky, O., and Bergen, J.: ‘Visual Odometry’, European Conference on Computer Vision (ECCV), 2004.