I_015

Optimization of 3D Egomotion Computation in Image Sequence

Yingdi Xie† Jun Ohya‡

1. Introduction

Egomotion estimation is one of the active researched topics among computer vision. It is a fundamental technology toward scene understanding in many applications, such as robot navigation, automated surveillance. In this paper, we proposed a method to compute the egomotion from time-sequential images acquired by active calibrated stereo cameras. This approach is developed to be implemented to the scene with independent motion. One of the challenges of this research is that independent motion and noise factor, such as wrong matching, will result in false egomotion computation. In other words, selecting the frame of reference is a tough task. To conquer this difficulty, previous literatures, such as M. Agrawal et al [1], utilize a random sample consensus (RANSAC, here-after) mechanism to evaluate a valid egomotion model. And most of them make the assumption that the egomotion with the maximal inlier number is the final result. Based on this assumption, the egomotion could be estimated effectively only when the independent moving object within scene occupies smaller region comparing to the background. Otherwise, however, the estimation will wrongly treat motion area as the background; continually result in false egomotion model. In this paper, we describe a method to compensate this deficiency according to the spatial distribution of the features, theoretically.

2. Egomotion model computation

In our work, egomotion is computed from optical flows within consecutive frames. By matching the feature points, which are detected by implanting Harris Corner Detector [5], numerous 2D optical flows including background flow, independent flow and wrong matching flow could be obtained simultaneously. Attach the depth information of each feature point by implementing matching between stereo images, optical flows within 3D space could be obtained. As described as follows, estimation of egomotion begins from randomly select three flows in disparity space. In our proposal, a method introduced by D. Demirdjian et al [2] is utilized. The egomotion model could be decomposed to two parts: translation and rotation, represented by the following equation:

$$M^i = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \quad (1)$$

where, R is a 3-by-3 matrix, and t is a 3-by-1 vector, denotes the rotation and translation factors respectively. Superscript i denotes the time instant of ith frame. The egomotion estimation is achieved by firstly computing rotation matrix. Translation vector is then computed by eliminating the motion generated from rotation.

2.1 Estimation from 3D optical flows

Egomotion estimation is computed by the following steps:

1) Denote the randomly selected flows with 3D coordinates. as $P_i(x_i,y_i,z_i)^T$, within the previous frame. $P'_i(x'_i,y'_i,z'_i)^T$ in the current frame. Where the subscript i ranges from 1 to 3, representing three different flows. Then estimate the rotation matrix R and the translation t such that $P'_i = R*P_i + t$.
2) Compute the centroid of the three points for each frame, denoted by:
$$\bar{P} = [\bar{x}_i, \bar{y}_i, \bar{z}_i]^T$$
$$\bar{P}' = [\bar{x}_i', \bar{y}_i', \bar{z}_i']^T \quad (2)$$
3) Subtract them with the original coordinates of each point. Then we have:
$$\hat{P}_i = [x_i - \bar{x} \quad y_i - \bar{y} \quad z_i - \bar{z}]^T$$
$$= [\hat{x}_i, \hat{y}_i, \hat{z}_i]^T$$
$$\hat{P}_i' = [\hat{x}'_i, \hat{y}'_i, \hat{z}'_i]^T \quad (3)$$
4) Compute the singular value decomposition (SVD) of $Q = USV^T$, where Q is given by:
$$Q = [\hat{P}_1, \hat{P}_2, \hat{P}_3] \begin{bmatrix} \hat{P}_1'^T \\ \hat{P}_2'^T \\ \hat{P}_3'^T \end{bmatrix} \quad (4)$$
5) Then rotation matrix $R = VU^T$ and translation vector is computed by $t = \bar{P}' - R\bar{P}$.

From the above steps, one egomotion could not be estimated by random selection of three flows. The validity for this model, however, could not be ensured. Because no preliminary knowledge indicates the 3D flows of selection composes of homogeneous motion area, especially belonging to background. Hence, in order to estimate an effective egomotion, RANSAC mechanism is adopted.

2.2 RANSAC mechanism and problem statement

Proposed by M. Fishler el at [3], RANSAC is one of the most effective regression methods applied to various researches. By briefly describing this method, it is a process that: Randomly selecting a minimum data set required to calculate a model, in this work, three optical flows in 3D space are randomly selected. Compute the egomotion model with the selected data set. For every optical flow within the original universal data set, count the inlier number whose distance is under a predefined distance threshold to the estimated model. Repeat the above step to select the model with the maximal number of inlier data, as the final result. From above description, by utilizing RANSAC, optical flows in 3D space are classified into two sets: background and non-background, including independent motion, error flows.

Obviously, the egomotion model is computed based on the assumption that most optical flows belong to the background. In reality, however, it is highly probable that the number of the optical flows that belong to the independent motions is larger than the one for the background. The above algorithm, hence, will lose its effectiveness against this case. Instead of selecting from inliers, we propose a method that optimizes the egomotion the egomotion model computation by the analysis of the spatial distribution of the optical flows.

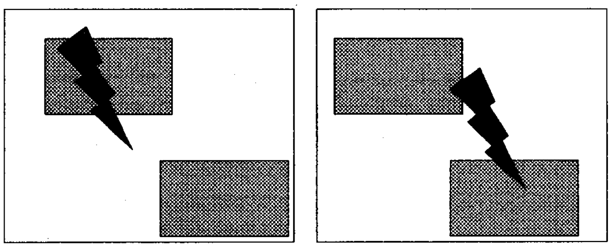
† Waseda Universtiy, GITS, Student

‡ Waseda University, GITS, Faculty

3. Optimization

By considering the ordinary scene, independent motion within it is always of consistency in spatial distribution. On the other hand, background may be segmented by multiple motion regions. According to these facts, we could assume that the computed homogeneous feature points with widest distribution belong to the background. Our method optimizes the egomotion computation by analyzing the spatial distribution of feature points. The distribution is testified by computing the variance of the coordinates of homogeneous feature point set VAR_{FP}. Thus, the RANSAC process is revised to:

1) Randomly select a minimal optical flow set to compute one egomotion model.
2) Compute the variance of the coordinates for feature points which are homogeneous to the egomotion model estimated above.
3) Repeat step 1) and 2), treat the homogeneous optical flow set as the background, whose variance value is the largest among all. Adopt the egomotion model computed from this flow set.

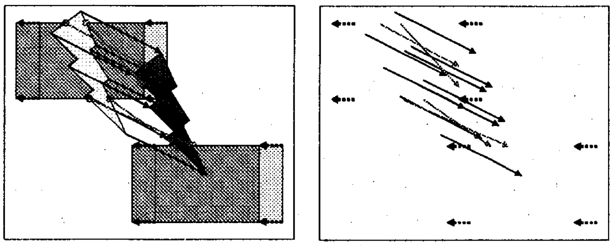


a. previous frame b. current frame

Fig. 1 Image sequence example. Two gray rectangles are background; the lighting shaped polygon is the independent moving object.

3.1 Example

Here, an example of distinguishing feature point by analyzing its spatial distribution is demonstrated. Refer to Fig.1.a, assume the background composes of two gray rectangles; the lightning shaped polygon is an independent moving object. In the next frame, independent motion area translating to the new position demonstrated as Fig.1.b. The camera is undergoing a right oriented translational egomotion. After applying feature detection by Harris Corner Detector, optical flows within two frames are then computed. Demonstrated in Fig.2.a, upper-left (yellow) area is undergoing an independent motion, to the middle lighting shaped region (blue) position in the current frame. Here, optical flows of the background denoted as dashed thick lines; optical flows of the independent motion denoted as real lines in black; red (gray) real lines represent for wrong detection and matching. From RANSAC process, two homogeneous frames of different motion areas could be constructed, while the wrong matching is eliminated. For the traditional criterion to recognize background, in this example, the optical flows that belong to the independent motions are falsely selected. Because the number of homogeneous flows for independent moving object has a larger amount than the one of background. By analyzing the spatial distribution, however, egomotion could be computed from the optical flows belong to background, because of their larger variance value.



a. overlapping two frames b. optical flows

Fig. 2 Optical flows belong to background has a wider spatial distribution. a. Upper-left lighting shaped area (yellow) is an independent motion moving toward the position of the middle lighting shaped area (blue). b. Optical flows, dashed lines are flows belong to the background, while real lines (black) belong to the motion region, and red real lines (gray) are error flows.

4. Discussion and Conclusion

The optimization of egomotion estimation could also be achieved from the image sequence. A typical method for this optimization is applying Extended Kalman Filter, mentioned in the work of D. Koller et al [4].

In this paper, we proposed a method to recover the egomotion in a more effective way than previous literatures. By reminding the traditional flow based egomotion estimation method, typically background features are recognized based when a homogeneous data set contains most features. In reality, however, it is probable that the number of the feature points belonging to the independent motions is larger than the one for the background. Thus, our proposal is to make the spatial distribution analysis of optical flows, since independent motion areas are usually within a certain region, in other words, they have a smaller distribution comparing to the background.

References

[1] M. Agrawal, K. Konolige, and L. Iocchi. Real-time detection of independent motion using stereo. In IEEE workshop on Motion (WACV/MOTION), January 2005.

[2] D. Demirdjian and T. Darrell. Motion estimation from disparity images. In *Proc. International Conference on Computer Vision*, volume 1, pages 213–218, July 2001.

[3] M. Fischler and R. Bolles. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. ACM.*, 24:381–395, 1981.

[4] D. Koller, G. Klinker, E. Rose, D. Breen, R. Whitaker, and M. Tuceryan, "Real-Time Vision-Based Camera Tracking for Augmented Reality Applications," Proc. ACM Symp. Virtual Reality Software and Technology (VRST-97), pp. 87-94, Sept. 1997.

[5] C.G. Harris and M. Stephens. A combined corner and edge detector. In fourth Alvey Vision Conference, pages 147 – 151, 1988