13—25

A Cooperative Method of SFM and Stereo for Motion and Depth Extraction

Jae-Hean Kim[1] Myung Jin Chung[2]
Department of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology

Abstract

Stereo is a useful technique for obtaining depth information from images. However, it is required that the baseline length between two cameras should be large to increase depth precision. Consequently, stereo matching suffers from ambiguities due to large baseline length. Therefore, the matching problem would become much easier if a sequence of densely sampled images along a camera path is used. "Structure from Motion (SFM)" do not suffer from the matching problem because many SFM algorithms use multiple images, especially a sequence of images taken at short time intervals. However, this method basically has one disadvantage of having scale factor ambiguity. The depth extraction method proposed in this paper integrates SFM and stereo depth extraction algorithm to determine a scale factor for SFM and to solve the matching problem of stereo even if there are ambiguities in matching and occlusion regions in the scene. Results on real images illustrate the performance of the proposed method.

1 Introduction

It is very important to recover the absolute depth of the scene in many applications of vision sensor based systems. Stereo is a useful technique for obtaining depth information from images. However, depth information from stereo is influenced seriously by matching problem. When the baseline length between cameras becomes longer to increase the precision of the results, the matching problem becomes more difficult [1]. Especially, it is certain when there are inherent ambiguities in matching, such as a repeated pattern over the large part of the scene and occlusion region which can not be seen from one of the camera. Therefore, the problem would become much easier if a sequence of densely sampled images along a camera path is used.

"Structure from Motion (SFM)" do not suffer from the matching problem because many SFM algorithms use multiple images, especially a sequence of images taken at short time intervals [2,3,4]. Also, there have been many researches about feature matching or tracking for SFM. Shi and Tomasi [6] proposed a feature selection criterion based on how tracker works and a feature monitoring method that can detect occlusions. However, this method basically has one disadvantage of having scale factor ambiguity. SFM based on feature tracking extracts basically the scene structure up to scale.

The depth extraction method proposed in this paper integrates SFM and stereo depth extraction algorithm to determine a scale factor for SFM and to solve the matching problem of stereo even if there are ambiguities in matching and occlusion regions in the scene. So, the proposed method is robust enough to be used in industry and society.

2 Cooperative Method

SFM and stereo in proposed method cooperate to remove the disadvantages of each algorithm. First, we extract the structure of the scene and motion parameters of the camera by using SFM. Then, we determine the scale factor by exploiting the reconstructed results for stereo matching process. Consequently, the proposed method reliably recovers the absolute range of the scene and motion parameters while the problem of scale factor ambiguity and matching problem of stereo are solved simultaneously.

Since the results from most SFM algorithm are up to scale, it constitute a set of structure that is spanned with respect to the origin of the camera coordinate system. So, the points of the set of structure can be reprojected on the other camera image plane that compose stereo as shown in Fig. 1. The stereo matching algorithm presented in this paper is adding the sum of squared difference (SSD) for all these projection points that are from the one of the elements of the structure set. Assume that SFM is performed on the left camera of stereo. Let us denote the N points on the scene by $\mathbf{M}_{Li}$, (i=1,...,N) which is reconstructed by SFM on some scale factor with respect to the left camera coordinate system and denote the rigid motion from left camera coordinate system to right camera coordinate system by rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$, then, the projected pixel positions $\mathbf{m}_{Ri}$, (i=1,...N) on the right camera plane for the reconstructed points are determined by (1) and (2)

$$\mathbf{M}_{Ri}(s)=\mathbf{R}(s\mathbf{M}_{Li})+\mathbf{t} \tag{1}$$

$$\tilde{\mathbf{m}}_{Ri}(s) = \mathbf{P}_R \tilde{\mathbf{M}}_{Ri}(s) \tag{2}$$

where, s is a scale factor and $\mathbf{P}_R$ is a perspective projection matrix of the right camera when world coordinate system is equal to the right camera coordinate system. Finally, we define a new evaluation function $e(s)$ for matching as follows

$$e(s) = \sum_{i=1}^{N} \sum_{\mathbf{d} \in W} (f(\mathbf{m}_{Li} + \mathbf{d}) - g(\mathbf{m}_{Ri}(s) + \mathbf{d}))^2 \tag{3}$$

where, $f(\mathbf{x})$ and $g(\mathbf{x})$ are image intensity function for the image of the left and right camera respectively, the $\Sigma_{\mathbf{d} \in W}$ means summation over a given feature window W and $\mathbf{m}_{Li}$, (i=1,...N) are pixel positions of interesting features on the left camera plane.

Then, we can extract absolute depths corresponding to interesting features and absolute camera motion parameters by searching scale factor s that minimizes $e(s)$ in (3).

[1] Address: 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701 Republic of Korea. E-mail: kjh@cheonji.kaist.ac.kr
[2] E-mail: mjchung@ee.kaist.ac.kr

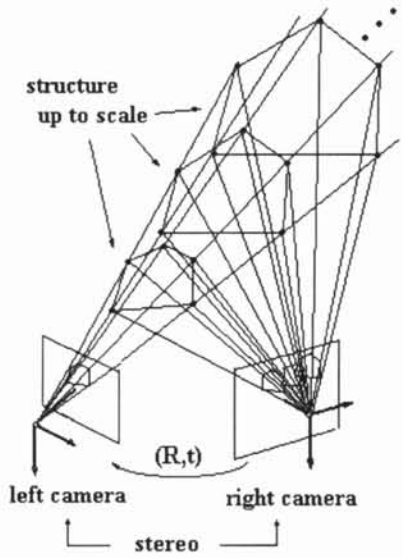


Fig.1. Reprojection of the scene points reconstructed by SFM on right camera when SFM is performed on left camera

3 Analysis of Cooperative Method

Owing to the proposed matching method using the evaluation function (3), we can overcome the problem of a false match even if there are inherent ambiguities in matching , such as a repeated pattern. In addition, if the number of non-occluded features are larger than that of occluded features (it is true in most case), the proposed method do not suffer from a false match even though some of the interesting features lie on occlusion regions in the scene. The advantage of the proposed method is depicted in Fig. 2. It is assumed that the intensity patterns around the feature 1,2,3 are same pattern. There is a great possibility of a false match if matching procedure is performed with individual feature as shown in Fig. 2 (a). Especially, since the feature 2 in occluding region cannot be seen from the right camera, it is impossible to find a true match. So, false match is to be extracted at all times for feature 2. However, if the proposed method is used as shown in Fig. 2 (b), false match would not occur. For example, even though the feature $\mathbf{m}_{L3}$ is matched with $\mathbf{m}_{R1}$ on a scale factor s_0, $\mathbf{m}_{L1}$ and $\mathbf{m}_{L2}$ would not be matched with $\mathbf{m'}_{R1}$ and $\mathbf{m'}_{R2}$ respectively. Consequently, evaluation function $e(s)$ in (3) would not have minimum value. On the other hand, when s_1 is selected as a scale factor, $e(s)$ would have minimum value because feature 1,3 serve true matches at $\mathbf{m}_{R1}$ and $\mathbf{m}_{R3}$ although $\mathbf{m}_{L2}$ do not match with $\mathbf{m''}_{R2}$. Fig. 2 (b) shows that the proposed method can give robust results against the repeated pattern and occlusion regions in the scene. In addition, if SFM is performed on the right camera and matching procedure is done in the left camera as well, we would acquire more plenty of information about depth of the scene.

4 Experimental Results

This section presents experimental results of the proposed method with real 2-D images. To track the features for SFM algorithm, we used Shi-Tomasi-Kanade tracker [5,6]. To extract structure of the scene up to scale, we use SFM algorithm proposed by Azarbayejani and Pentland [2], which has shown good performance in our experiments. In this experiment 100 images of sequence

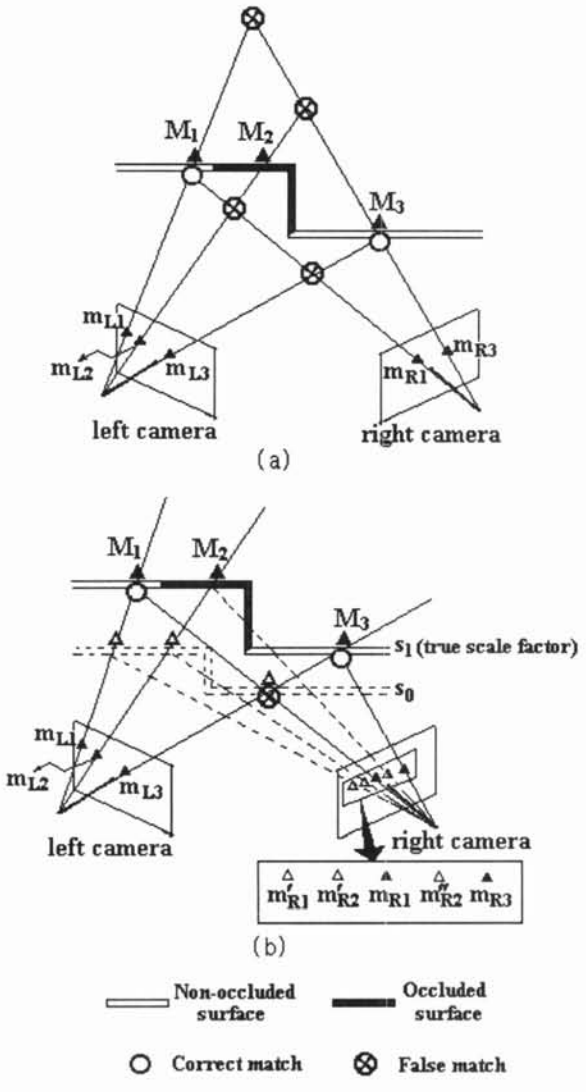


Fig. 2. (a) Matching with individual feature matching. (b) Matching with proposed method.

are used. The sequence was acquired by manually moving the stereo. The size of used images is 640×480 pixels and window size for matching is 3×3 pixels.

Fig. 3 shows image data set used for experiment. This image data set has many ambiguities in matching because of repeated patterns over the large part of the scene and occlusion regions due to foreground pillar. In Fig. 3, the first and last frames of the stereo sequence are shown. Feature points and tracking results are shown in left image of Fig. 3 (a) and (b) respectively.

Reconstructed structure and left camera pose from SFM are shown in Fig. 4. Fig. 5 shows the matching result by using the proposed method about the first stereo image frames. From these results, we can observe that the proposed method can give robust matching results. In Fig. 6, evaluation function $e(s)$ for this experiment is shown. Fig. 7 shows results of false match for some features when matching procedure is performed with individual feature through the epipolar line. By means of this matching result, we can determine the scale factor for the structure and motion parameters from SFM. Table 1 shows the quantitative results. Search step to find scale factor is 0.1mm.

	Scale factor (mm)
Ground truth value	321.5
Estimated value	321.2

Table. 1. Accuracy of the results

References

[1] U. R. Dhond and J. K. Aggarwal, "Structure from stereo - a review," IEEE Transaction on Systems, Man, and Cybernetics, vol. 19, no. 6, pp.1489-1510, 1989.

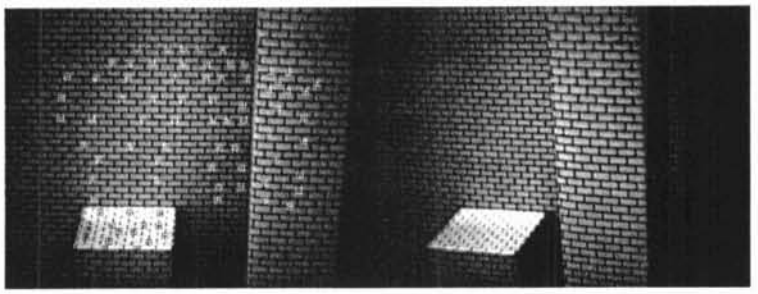
[2] A. Azarbayejani and A. Pentland, "Recursive estimation of motion, structure, and focal length," IEEE Transaction on Pattern Recognition and Machine Intelligence, vol. 17, no. 6, pp.562-575, June. 1995.

[3] R. Szeliski and S. Kang, "Recovering 3d shape and motion from image streams using nonlinear least squares," Journal of Visual Communication and Image Representation, vol. 5, no. 1, pp10-28, 1994.

[4] T.J. Broida, S.Chandrashekhar, and R.Chellapa, "Recursive estimation of 3D motion from a monocular image sequence," IEEE Transaction on Aerospace and Electronics Systems, vol. 26, no. 4, pp.639-656, July. 1990.

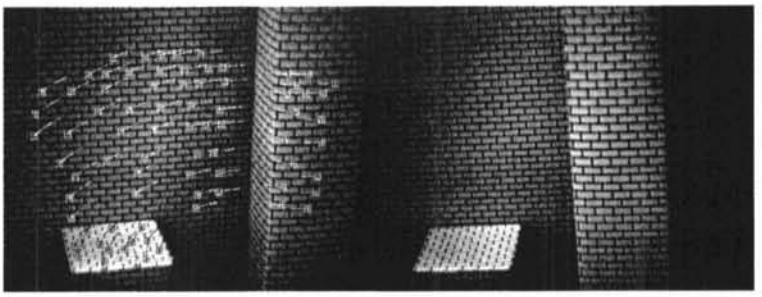
[5] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," In Int. Joint Conference on Artificial Intelligence, pp.121-130, Aug. 1981.

[6] J. Shi and C. Tomasi, "Good features to track," In IEEE Conference on Computer Vision and Pattern Recognition, pp.593-600, June. 1994.



Left image Right image

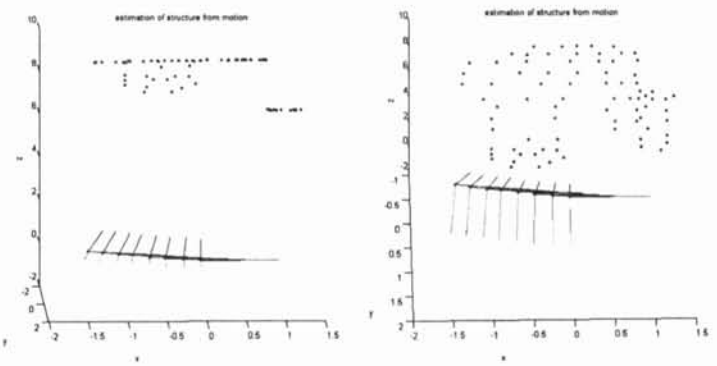
(a)



Left image Right image

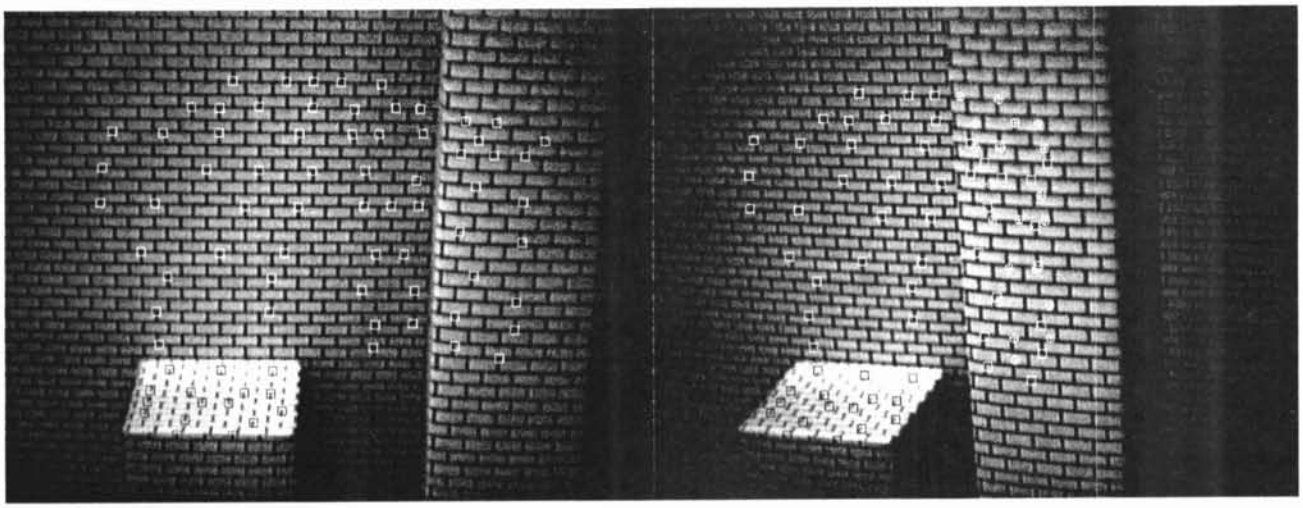
(b)

Fig. 3. First (a) and last (b) frames of the stereo sequence: Feature points (left image in (a)) ; tracking results (left image in (b)). There are many repeated patterns over the large part of the scene and occlusion regions due to the foreground pillar.



(a) (b)

Fig.4. Reconstructed structure and left camera pose from SFM on some scale factor: Top view (a) and side view (b)



Left image Right image

Fig. 5. Matching results with the proposed method: The feature points behind the pillar can not be seen in right image but the virtual positions indicated by " ⊕" are acquired due to the proposed method.

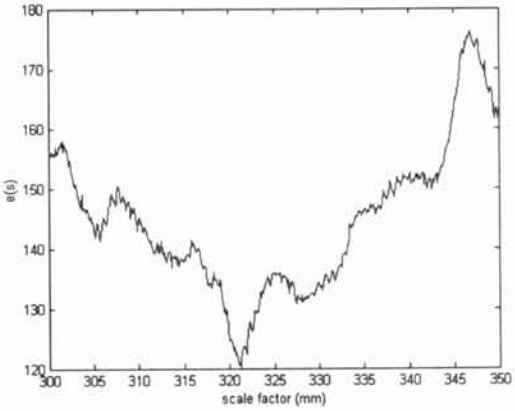
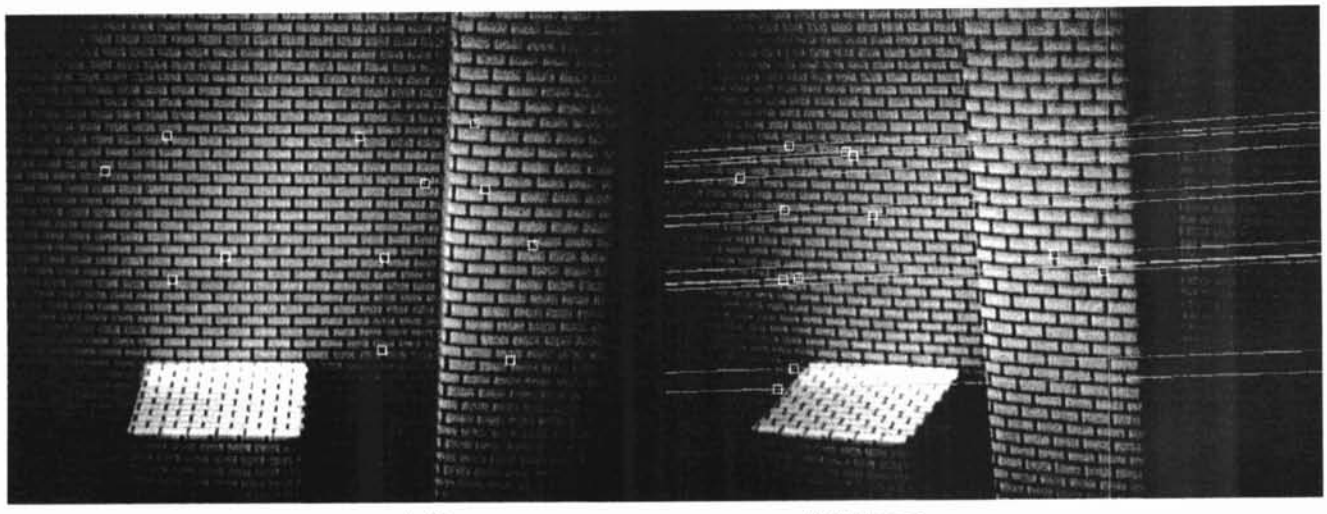


Fig. 6. Value of evaluation function



Left image Right image

Fig. 7. Matching result for some features with individual feature matching through the epipolar line: False matching due to the repeated pattern and occlusion are observed.