Sequential Motion and Scene Reconstruction from Image Sequences Captured by a Multi-camera System

Jae-Hean Kim, Myung Jin Chung [†], Chang Joon Park, and In Ho Lee
Digital Content Research Division
ETRI, Republic of Korea
[†] Department of Electrical Engineering and Computer Science
KAIST, Republic of Korea
{gokjh,chjpark,leeinho}@etri.re.kr, [†]mjchung@ee.kaist.ac.kr

Abstract

This paper deals with structure and motion estimation by using image sequences captured by a multi-camera system moving in a scene. Single camera systems with a relatively small field of view have limited accuracy because of the inherent confusion between translation and rotation. This is also the case for the stereo camera systems if the systems need the intersection of the fields of view for stereo correspondence. The cameras constituting the multi-camera system considered ultimately in this paper are arranged so that there is no, or a small, intersection of the fields of view. This configuration improves the accuracy of the structure and motion estimation because the ambiguity mentioned above decreases due to a large field of view. We propose a recursive algorithm for real time applications of the multi-camera system using long term image sequences, which have appearing and disappearing point features. The results of the experiments with long term real image sequences are presented to demonstrate the feasibility of the algorithm.

1 Introduction

There have been many attempts to reconstruct both the motion of cameras and the structure of a scene simultaneously. This research is known as *structure from motion* (SFM), which has been an important topic in computer vision and has been approached in a variety of ways. However, there are some issues to be argued for a SFM algorithm to be applied to the robotic system.

First, the main issue is the problem of the limited field of view. Most of the work in the SFM has concentrated on single camera systems. However, the SFM algorithm using a single camera has a limited accuracy due to a relatively small field of view. The reason is that there is an inherent ambiguity of confusion between translation and rotation [6]. This effect is accumulated during long term camera movement and can be a significant factor of motion drift. This is also the case for the stereo camera systems moving in a scene if the systems need the intersection of the fields of view for stereo correspondence.

To remove this ambiguity, several approaches have been suggested. These approaches can be categorized into three classes. First, a catadioptric vision sensor, which consists of a mirror and a camera, can widen the camera's field of view [11]. However, the sensor system has the disadvantage of non-uniform and low angular resolution, which can affect the estimation accuracy. Second, an active vision system that can change the viewing direction and see a wide angular range can be an alternative method [5]. The shortcomings of this system are a cumbersome mechanism and a finite time required to capture an image with a wide angular range. Finally, we can adopt a multi-camera system [10], [2], [9]. The cameras constituting a multi-camera system are arranged so that there is no or small intersection of the fields of view. Consequently, multi-camera system can acquire a large field of view and alleviate the ambiguity of confusion between translation and rotation, keeping uniform and high image resolution. The advantage of using multi-camera system and the design strategy are well discussed in [2] and [9]. Since, in these days, fully digital cameras being able to acquire images with high image resolution and transferring them to PC at video rate are available inexpensively, it will be more convenient and useful to install a multi-camera system.

Another issue to be discussed is the *scale factor ambiguity*, which is an inherent ambiguity of monocular motion vision. This ambiguity can cause a scale factor drift during long term camera movement [3]. It is worthwhile to note that an absolute scale factor can be acquired while a stereo camera moves in a scene without feature correspondences among the cameras. This result was proved in [7]. These results imply that a multi-camera system of multi-viewpoint structure can obtain absolute estimates although there is no intersection of the fields of view to acquire wider field of view.

The issue we want to mention finally is the real-time requirement. To acquire current estimates robust to image noise, it is necessary to integrate all visual information up to current time and, in robotic applications, each new image must be processed in a short time to produce estimates as it becomes available. This real-time requirement has led to the development of an on-line method, which generally has a recursive form using the *extended Kalman filter* (EKF) [1], [3], [4]. Besides the real-time requirement, it is essential during long term camera movement for practical SFM algorithm to handle point features appearing and disappearing due to the limited field of view and the tracking problems, for instance, occlusion, disocclusion, specularity, etc.

However, the real-time SFM algorithm satisfying above all requirements has not been well discussed for multi-camera systems. Motion estimation algorithms for multi-camera system in [2], and [9] were designed to obtain differential motion estimates or for a two-view framework. Recently, an approach acquiring the camera motion parameter by using a multi-camera system for long term image sequences was suggested by Sato *et al.* [10]. However, this method is a batch method and requires some features of known 3-D positions. The methods using a stereo camera system have shown good results for motion and structure estimation, which have an absolute scale factor [8]. However, as long as the systems need stereo correspondence, there is a limit for widening the field of view because a scene should be in the intersection of the fields of view.

In this paper, we propose a recursive SFM algorithm for real time applications of multi-camera system using long term image sequences, which have point features appearing and disappearing. Our approach can be viewed as an extension of the work by Azarbayejani *et al.* [1] and Chiuso *et al.* [3] to a multi-camera system considering scale factor determinacy. We found in experiments that using just two cameras can obtain much
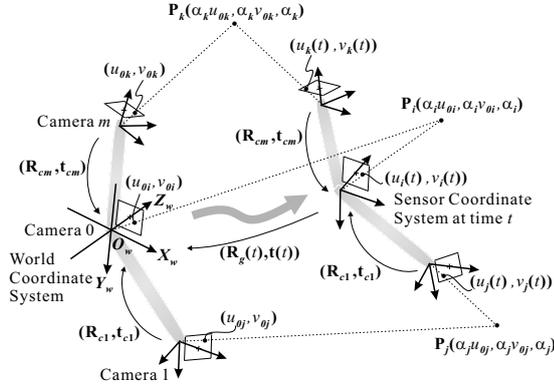
Figure 1: A schematic diagram of the multi-camera system in the process of observing features in each camera.

better accuracy than using a single camera. We also extend the proof of the degenerate motions in determining an absolute scale factor without correspondence among the cameras from two cameras to a multi-camera system, which has more than two cameras. The proposed method assumes that the intrinsic parameters of cameras and the geometries between cameras are given and the scene viewed from the cameras is static.

2 Realization of the algorithm

2.1 State of the EKF

Fig. 1 shows a schematic diagram of the multi-camera system in the process of observing features in each camera. From now on, we assume that one of the camera coordinate systems is selected as the sensor coordinate system and define this camera coordinate system as a *reference camera coordinate system*. The six motion parameters of the sensor are represented with $\mathbf{R}_g(t)$ and $\mathbf{t}(t)$.

The position of the ith feature in 3D is parameterized as

$$\mathbf{P}_i = [\alpha_i u_{0i}, \alpha_i v_{0i}, \alpha_i]^T, \tag{1}$$

where α_i and (u_{0i}, v_{0i}) are the depth and normalized image coordinates of the ith feature respectively with respect to each camera coordinate system of the start position.

The state vector for the EKF consists of six motion parameters, six velocity parameters, one baseline length parameter, and $3N$ structure parameters, where N is the number of features tracked.

$$\mathbf{x}(t) = \left[\mathbf{t}^T(t), \rho^T(t), \dot{\mathbf{t}}^T(t), \dot{\rho}^T(t), b_l, \alpha_1, u_{01}, v_{01}, \cdots, \alpha_N, u_{0N}, v_{0N}\right]^T \tag{2}$$

The rotational vector $\rho(t)$ consists of the XYZ Euler angles that describe the small interframe rotational motion. This interframe rotation $\rho(t)$ is estimated at each frame and then composed with the global rotation represented by $\mathbf{R}_g(t)$ as follows.

$$\mathbf{R}_g(t) = \mathbf{R}(\rho(t))\mathbf{R}_g(t - \Delta t), \tag{3}$$

where $\mathbf{R}(\rho(t))$ means the rotation matrix generated by the XYZ Euler angle vector $\rho(t)$. Due to this parameterization, the angles will remain small and the singularity associated with the Euler angle representation can be avoided because the singularity occurs when the second rotation angle is $90°$. The reader who is uncomfortable with modelling rotation as the Euler angle may consider angle-axis or quaternion parameterizations.

The b_l is the baseline length between reference camera and one of the other cameras. Although it is assumed that the baseline lengths are given by the geometries between the cameras, the state vector includes one baseline length to improve the convergence rate. The reason is explained in detail in [7].

2.2 Measurement Model

As shown in Fig. 1, the position of the ith feature point observed in the mth camera can be represented with respect to the current mth camera coordinate system via

$$\begin{bmatrix} P_{xi}^{cur} \\ P_{yi}^{cur} \\ P_{zi}^{cur} \end{bmatrix} = \mathbf{R}_{c_m} \left[\mathbf{R}_g(t) \mathbf{R}_{c_m}^T \left(\mathbf{P}_i - \mathbf{t}_{c_m} \right) + \mathbf{t}(t) \right] + \mathbf{t}_{c_m}, \tag{4}$$

where $\mathbf{R}_{c_m}$ and $\mathbf{t}_{c_m}$ is the rotation matrix and translation vector of the reference camera coordinate system respectively relative to the mth camera coordinate system. For reference camera, $m = 0$ and $\mathbf{R}_{c_0}$ and $\mathbf{t}_{c_0}$ are the 3×3 identity matrix and the 3×1 zero vector respectively.

Referring to Eqs. (1), (3) and (4) together, a non-linear measurement model for the ith feature on a normalized image plane can be written as

$$\begin{aligned} \mathbf{y}_i(t) &= [u_i, v_i]^T + \eta_i(t) \\ &= \left[\frac{P_{xi}^{cur}}{P_{zi}^{cur}}, \frac{P_{yi}^{cur}}{P_{zi}^{cur}} \right]^T + \eta_i(t) \\ &= \mathbf{g}_i(\mathbf{t}(t), \rho(t), b_l, \alpha_i, u_{0i}, v_{0i}, \mathbf{R}_g(t - \Delta t), \mathbf{R}_{c_m}, \mathbf{a}_{c_m}, r_m) + \eta_i(t) \end{aligned} \tag{5}$$

relating the structure of a feature and the motion of a sensor to the current normalized image coordinates observing the feature, where $\eta_i(t)$ is the noise term associated with the measurements of image coordinates and modelled as a zero-mean Gaussian white noise sequence and $\mathbf{a}_{c_m}$ is the unit vector of $\mathbf{t}_{c_m}$ and r_m is the ratio of b_l to the baseline length between the reference camera and the mth camera. Accordingly, $\mathbf{t}_{c_m} = r_m b_l \mathbf{a}_{c_m}$.

The overall measurement model is a set of N 2D measurements as follows:

$$\begin{aligned} [\mathbf{y}_1(t)^T, \dots, \mathbf{y}_N(t)^T]^T = \mathbf{g}(\mathbf{x}(t), \mathbf{R}_g(t - \Delta t), \mathbf{R}_{c_1}, \mathbf{a}_{c_1}, r_1, \cdots, \mathbf{R}_{c_M}, \mathbf{a}_{c_M}, r_M) \\ + [\eta_1(t)^T, \dots, \eta_N(t)^T]^T \end{aligned} \tag{6}$$

where M is the number of cameras. The non-linear vector function $\mathbf{g}(\cdot)$ of Eq. (6) is the vertical accumulation of the function $\mathbf{g}_i(\cdot)$ of Eq. (5).

2.3 Dynamic Model

The dynamic model for the EKF is the linear dynamic model presented in [7]. In that work, it is assumed that the sensor's velocity varies very slowly. This assumption is feasible due to inertia, gravity, and other physical constraints.

The details of the recursion to update the state and the state covariance matrix of the EKF are not developed further here because they are straightforward.

2.4 Scale Factor Determinacy

It was proved in Theorem 2 of [7] that the scale factor can be determined if, and only if, the motion is not included in a degenerate motion set. From the Theorem 2, we can easily deduce the degenerate cases for a multi-camera system, which has more than two cameras and fixed geometries among the cameras. Let D_2^1, D_2^2, and D_2^3 be the three motion sets in order described in Theorem 2 for one of the camera pairs in a multi-camera system and let D_m be the degenerate motion set for a multi-camera system. Also, let $D_2 = D_2^1 \cup D_2^2 \cup D_2^3$.

Corollary 1 *Assume that the motion parameters of the cameras in multi-camera system are given up to scale and the geometries among the cameras are known. It is also assumed that all of the cameras have non-zero translational motion. If the baselines of every camera pairs are*

  1. *not on a same line, then $D_m = D_2^1$.*

  2. *on a same line, then $D_m = D_2$.*

proof It can be seen that $D_2^1 \cap (D_2^2 \cup D_2^3) = \phi$ and $D_m \supset D_2^1$ from Theorem 2. It is also clear that $D_2 \supset D_m$. If the baselines of every camera pairs are

1) not on a same line, it can be readily established that $D_m \not\supset (D_2^2 \cup D_2^3)$ by using Theorem 2. Then, $D_2^1 = D_m$.

2) on a same line, it can also be established that $D_m \supset (D_2^2 \cup D_2^3)$ by using Theorem 2. Then, $D_2 = D_m$. $\square$

From the Corollary 1, we can see that, if the baselines of every camera pairs are not on a same line, the degeneracy only occur during the pure translational motion of multi-camera system having more than two cameras. Moreover, regardless of the number of cameras, all the degenerate motions can be avoided with the help of a simple 1D rotational mechanism on which the multi-camera system can be mounted, and the angle of the rotational motion need not be so large to remove the degeneracy in the proposed method [7].

The scale factor can be determined from the constraints derived from the geometries between the cameras. The constraints are enforced in Eq. (6) because Eq. (4) implies the relationship between the motion of the reference and the other cameras. Therefore, the EKF using the measurement model of Eq. (6) can compute the absolute structure and motion estimates.

2.5 Removal of Point Features Disappearing

When a feature disappears or is mistracked, the corresponding state elements should be removed from the EKF process and, if the variance is under some threshold, its best current estimates are kept in the memory vector to store past structure data. If ith feature disappears, we eliminate the elements of the state vector, $\{\alpha_i, u_{0i}, v_{0i}\}$, and the rows and the columns of $\hat{\mathbf{P}}_\mathbf{x}$ corresponding to these states. In addition, since we cannot acquire the measurement of that feature anymore, the measurement model corresponding to that feature is eliminated from Eq. (6).

2.6 Initialization of New Point Features Appearing

Whenever a feature is removed, a new point feature is extracted from the current image frame to maintain a roughly constant number of features determined at the start of the algorithm. When a new feature is extracted, it must be inserted into the state of the EKF. However, since the initial condition of the states is unknown, especially for depth information, if we insert it simply modelled with a large variances into the $\hat{\mathbf{P}}_\mathbf{x}$, the filter goes through an unwanted transient. During this transient, error propagation affects the estimation accuracy of the other states and even may cause the filter to diverge due to the linearization characteristics of the EKF. To address this problem, we run the additional EKF for each new feature determining reasonably the initial values for the states. This filter estimates the position of the feature by using the motion, $\{\hat{\mathbf{t}}, \hat{\rho}\}$, and the baseline length, $\{\hat{b}_l\}$, estimated currently by the existing main filter.

Suppose that the jth feature is a new extracted feature observed in the mth camera at time τ and $\mathbf{R}_{r_j} = \mathbf{R}_g(\tau)$ and $\mathbf{t}_{r_j} = \mathbf{t}(\tau)$. We save the $\{\mathbf{R}_{r_j}, \mathbf{t}_{r_j}\}$ in the memory for the new features appearing at time τ. The position of the feature point can be represented with respect to the current mth camera coordinate system via

$$\begin{bmatrix} P_{xj}^{cur} \\ P_{yj}^{cur} \\ P_{zj}^{cur} \end{bmatrix} = \mathbf{R}_{c_m} \left\{ \mathbf{R}_g(t)\mathbf{R}_{r_j}^T \left[\mathbf{R}_{c_m}^T \left(\mathbf{P}_j - \mathbf{t}_{c_m} \right) - \mathbf{t}_{r_j} \right] + \mathbf{t}(t) \right\} + \mathbf{t}_{c_m}, \tag{7}$$

where $\mathbf{P}_j = \left[\alpha_j u_{0j}, \alpha_j v_{0j}, \alpha_j \right]^T$, in which α_j and (u_{0j}, v_{0j}) are the depth and normalized image coordinates of the jth feature respectively with respect to the mth camera coordinate system at time τ. From Eq. (7), the measurement model of the new feature on a normalized image plane can be written as

$$\begin{aligned} \mathbf{y}_j(t) &= \left. [u_j, v_j]^T \right|_{(\hat{\mathbf{t}}, \hat{\rho}, \hat{b}_l)} + \eta_j(t) \\ &= \left. \begin{bmatrix} \dfrac{P_{xj}^{cur}}{P_{zj}^{cur}}, \dfrac{P_{yj}^{cur}}{P_{zj}^{cur}} \end{bmatrix}^T \right|_{(\hat{\mathbf{t}}, \hat{\rho}, \hat{b}_l)} + \eta_j(t) \\ &= \mathbf{h}_j(\hat{\mathbf{t}}, \hat{\rho}, \hat{b}_l, \alpha_j, u_{0j}, v_{0j}, \mathbf{R}_g(t - \Delta t), \mathbf{R}_{c_m}, \mathbf{a}_{c_m}, r_m, \mathbf{R}_{r_j}, \mathbf{t}_{r_j}) + \eta_j(t), \end{aligned} \tag{8}$$

where $\eta_j(t)$ is the same as defined in Eq. (5). It is worthwhile to note that the state vector to be estimated in $\mathbf{h}_j(\cdot)$ of Eq. (8) is only

$$\mathbf{w}_j(t) = [\alpha_j, u_{0j}, v_{0j}]^T. \tag{9}$$

The linear dynamic model is

$$\begin{cases} \alpha_j(t + \Delta t) = \alpha_j(t) \\ u_{0j}(t + \Delta t) = u_{0j}(t) \\ v_{0j}(t + \Delta t) = v_{0j}(t). \end{cases} \tag{10}$$

By using the models (8) and (10), the additional filter estimates the position of the jth feature. After the convergence of the state of the feature, especially for the $\{\alpha_j\}$, it can be inserted into the state of the main filter. The criterion we use to decide when the state converges enough is when the variance of the $\{\alpha_j\}$ is less than a given threshold.

Suppose that the jth feature converges. It is very simple to insert the state into the state of the main filter. The measurement model for the new jth feature in the main filter is

$$
\begin{aligned}
\mathbf{y}_j(t) = \ & \mathbf{h}_j(\mathbf{t}(t), \boldsymbol{\rho}(t), b_l, \alpha_j, u_{0j}, v_{0j}, \mathbf{R}_g(t - \Delta t), \\
& \mathbf{R}_{c_m}, \mathbf{a}_{c_m}, r_m, \mathbf{R}_{r_j}, \mathbf{t}_{r_j}) + \eta_j(t).
\end{aligned} \tag{11}
$$

Note that $\mathbf{h}_j(\cdot)$ of Eq. (11) and of Eq. (8) are the same functions but $\{\mathbf{t}(t), \boldsymbol{\rho}(t), b_l\}$ in Eq. (11) are states to be estimated. The linear dynamic model for the new feature in the main filter is the same as model (10). What we must do is only to append model (10) to the linear dynamic model of the main filter and model (11) to model (6). Then, the additional filter for the jth feature is removed and the state vector and state covariance matrix estimated up to now by the additional filter corresponding to the jth feature are inserted in to the main filter.

During the first transient period of the main filter when the motion states are not yet converged, new features are not extracted although there is a feature disappearing. This is because the initial depth estimation of the new feature in the additional filter explicitly depends on the motion and baseline states estimated from the main filter (see Eq. (8)). We test the transient by using both a threshold on the variance of the depth estimates and a given time interval.

3   The Issues of Initial Conditions

The initial conditions for the main filter are as follows.

$$
\begin{cases}
[\mathbf{t}(0), \boldsymbol{\rho}(0), \dot{\mathbf{t}}(0), \dot{\boldsymbol{\rho}}(0)] = [\mathbf{0}_{3\times1}, \mathbf{0}_{3\times1}, \mathbf{0}_{3\times1}, \mathbf{0}_{3\times1}] \\
b_l(0) = \nu \\
[\alpha_i(0), u_{0i}(0), v_{0i}(0)] = [1, \tilde{u}_i^0, \tilde{v}_i^0] \qquad i = 1, \cdots, N,
\end{cases}
$$

where ν is selected typically over $[0.01, 0.1]$, $\{\tilde{u}_i^0, \tilde{v}_i^0\}$ is the actual measurement of the image coordinates of the ith feature at the start position, and the sensor coordinate system at the start position is assumed to be a world coordinate system. Although the initial value for the $\{b_l\}$ is somewhat arbitrary and may cause a large scale factor error, we can acquire fast convergence rate due to the parameterization of the baseline length.

The initial state covariance matrix for the main filter is as follows.

$$
\mathbf{P_x}(0) = \Delta_{diag}\left(\mathbf{0}_{3\times3}, \mathbf{0}_{3\times3}, \mathbf{0}_{3\times3}, \mathbf{0}_{3\times3}, \lambda_{b_l}, \lambda_{\alpha_1}, \mathbf{R}_1^0, \cdots, \lambda_{\alpha_N}, \mathbf{R}_N^0\right), \tag{12}
$$

where $\Delta_{diag}(\cdot)$ means that a block diagonal matrix with given arguments and λ_{b_l} and λ_{α_i}, for $i = 1 \cdots, N$, are selected typically over $[100, 1000]$ and $[0.1, 1]$ respectively. Since the absolute scale factor is completely unknown from the initial frame, we assign a large positive number to λ_{b_l}.

The initial conditions for the additional filter are as follows.

$$
[\alpha_j(\tau), u_{0j}(\tau), v_{0j}(\tau)] = [1, \tilde{u}_j^\tau, \tilde{v}_j^\tau], \tag{13}
$$

where $\{\tilde{u}_j^\tau, \tilde{v}_j^\tau\}$ is the actual measurement of the image coordinates of the new jth feature when the feature appears at time τ.

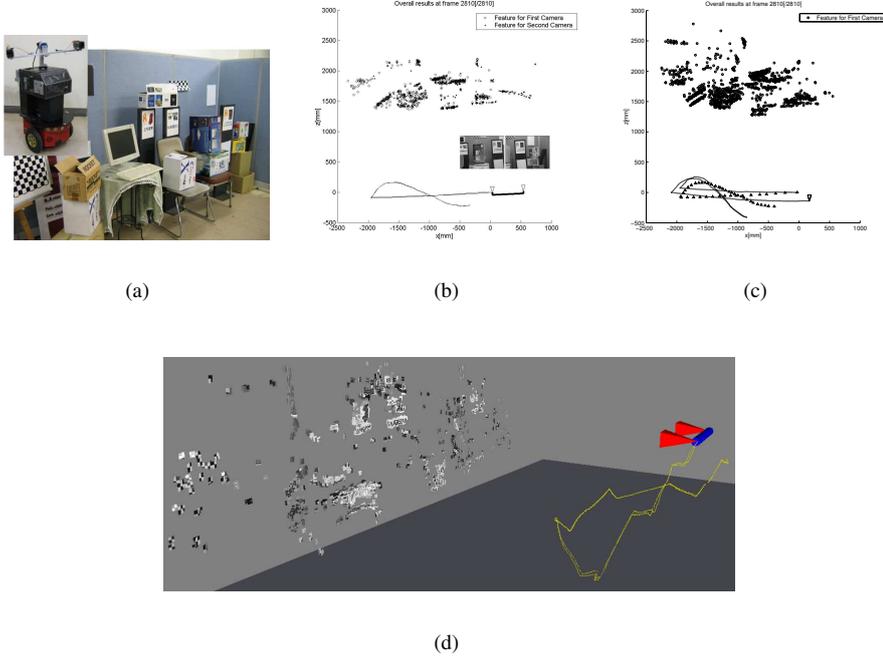(a)   (b)   (c)

(d)

Figure 2: (a) A two-camera system attached on a mobile robot and the side view of the experimental environment are shown. The top view of the overall estimation results for (b) the proposed method and (c) the single camera method. (d) The graphic representation of the estimation results.

The initial state covariance matrix for the additional filter is as follows.

$$\mathbf{P}_{[\alpha_j, u_{0j}, v_{0j}]}(\tau) = \Delta_{diag}\left(\lambda_{\alpha_j}^{new}, \mathbf{R}_j^{\tau}\right), \tag{14}$$

where $\lambda_{\alpha_j}^{new}$ is typically 1000 and $\mathbf{R}_j^{\tau} = E[\eta_j(\tau)\eta_j^T(\tau)]$. The initial variance for the new feature, $\lambda_{\alpha_j}^{new}$, is much larger than λ_{α_i}, which is for the feature observed from the start position. This is due to the fact that there is an unexpected and larger range of depths of features in long term image sequences with large movements.

4   Experimental Results

In this section, experimental results of applying the proposed SFM method to real image sequences are presented. The real image sequences were obtained by using a camera system attached on a mobile robot, which was driven in office environment. The camera system was mounted on a 1D rotational mechanism that could have roll motion to avoid the degenerate motion. The experimental environment is shown in Fig. 2(a).

Fig. 2(b) and 3(a) show the overall estimation results. The last image frames are also shown. The circles and dots represent the reconstructed feature positions and the solid
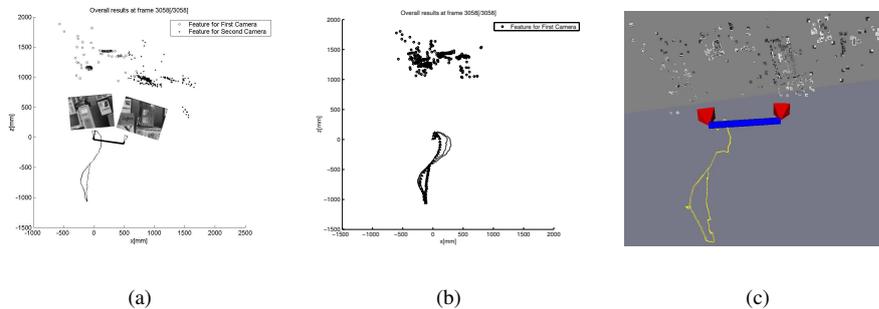
Figure 3: The top view of the overall estimation results for (a) the proposed method and (b) the single camera method. (c) The graphic representation of the estimation results.

line is the estimated path of the multi-camera system. The pyramid represents the camera viewing volume at the last frame. Fig. 2(d) and 3(c) show the graphic representation of the reconstructed structure and the camera trajectory. Textures of the feature image patches are mapped on the reconstructed position. The pyramid represents the camera viewing volume. The overall image sequences were 1405 and 1529 frames long and were repeated backward as if the camera system was moved back to the initial position. Then, we measured the differences between the forward and backward path estimated. Since there is no ground truth in this experiment, these differences will be a good index of the accuracy of the motion estimates. The difference of translational motion at the last frame is about 35.07mm and 20.96mm, respectively.

We also present the experimental results of the SFM of using single camera for the same image sequences. We refer to this method as *single camera method*. The purpose of this experiment was to compare these results with the results of the proposed method of using multi-camera system. Fig. 2(c) and 3(b) shows the overall estimation results of the single camera method for the first camera of the two cameras. The motion estimates by the proposed method, which is indicated by "△" is also shown for comparison. We can explicitly see the difference between the results of the two methods. We can see that the difference of the moving length between the forward and backward path, which shows some evidence to the scale factor drift.

5   Conclusion

The purpose of the proposed method is to overcome the limitations of the previous approaches that estimate both the motion of cameras and the structure of a scene as the camera system moves through it. The main limitation is the field of view of the cameras. The limited field of view causes confusion between translation and rotation and affects the accuracy of the algorithm that uses a single or stereo camera system. We suggest that implementing an SFM algorithm using a multi-camera system is reasonable to remove this ambiguity of confusion. The proposed algorithm was formulated to be a recursive form for real-time applications and can handle point features appearing and disappearing. Moreover, while the proposed algorithm can widen the field of view of the system by

arranging the cameras to view different part of the scene and, therefore, does not need the stereo correspondence, it does not degrade the advantage of the stereo camera system, that it can acquire absolute estimates. The proposed method has been tested in the experiments with long term real image sequences to demonstrate the feasibility of the algorithm.

References

[1] Ali Azarbayejani and Alex P. Pentland. Recursive estimation of motion, structure, and focal length. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(6):562–575, June 1995.

[2] Patrick Baker, Abhijit S. ogale, and Cornelia Fermüller. The argus eye: a new imaging system designed to facilitate robotic tasks of motion. *IEEE Robot. Automat. Mag.*, 11(4):31–38, December 2004.

[3] Alessandro Chiuso, Paolo Favaro, Hailin Jin, and Stefano Soatto. Structure from motion causally integrated over time. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(4):523–535, April 2002.

[4] Andrew J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proc. IEEE International Conference on Computer Vision*, pages 1403–1410, Beijing, China, October 2003.

[5] Andrew J. Davison and David W. Murray. Simultaneous localization and map-building using active vision. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(7):865–880, July 2002.

[6] Cornelia Fermüller and Yiannis Aloimonos. Observability of 3D motion. *International Journal of Computer Vision*, 37(1):43–63, June 2000.

[7] Jae-Hean Kim and Myung Jin Chung. Recursive estimation of absolute stereo motion and structure without stereo correspondence. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3387–3393, Sendai, Japan, September-October 2004.

[8] David Nistér, Oleg Naroditsky, and James Bergen. Visula odometry. In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pages 652–659, Washington, DC, USA, June-July 2004.

[9] Robert Pless. Camera cluster in motion: motion estimation for generalized camera designs. *IEEE Robot. Automat. Mag.*, 11(4):39–44, December 2004.

[10] Tomokazu Sato, Sei Ikeda, and Naokazu Yokoya. Extrinsic camera parameter recovery from multiple image sequences captured by an omni-directional multi-camera system. In *Proc. European Conference on Computer Vision*, pages 326–340, May 2004.

[11] Dennis Strelow, Jeff Mishler, Sanjiv Singh, and Herman Herman. Extending shape-from-motion to noncentral omnidirectional cameras. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2086–2092, Maui, Hawaii, USA, October-November 2001.