Extrinsic Camera Parameter Recovery from Multiple Image Sequences Captured by an Omni-directional Multi-camera System

Tomokazu Sato, Sei Ikeda, and Naokazu Yokoya

Graduate School of Information Science,
Nara Institute of Science and Technology,
8916-5 Takayama, Ikoma, Nara 630-0192, Japan,
{tomoka-s,sei-i,yokoya}@is.aist-nara.ac.jp ,
WWW home page: http://yokoya.aist-nara.ac.jp/

Abstract. Recently, many types of omni-directional cameras have been developed and attracted much attention in a number of different fields. Especially, the multi-camera type of omni-directional camera has advantages of high-resolution and almost uniform resolution for any direction of view. In this paper, an extrinsic camera parameter recovery method for a moving omni-directional multi-camera system (OMS) is proposed. First, we discuss a perspective n-point (PnP) problem for an OMS, and then describe a practical method for estimating extrinsic camera parameters from multiple image sequences obtained by an OMS. The proposed method is based on using the shape-from-motion and the PnP techniques.

1 Introduction

In recent years, many types of omni-directional cameras have been developed [1–6] and have attracted much attention in a number of different fields such as robot navigation, telepresence and video surveillance. Especially, the omni-directional multi-camera system (OMS) [4–6] has some advantages in the field of augmented virtuality because that provides high-resolution and almost uniform resolution for any direction of view. Generally, an OMS has multiple camera units that are located radially and are fixed in certain positions in the camera block of the OMS. Although an OMS has many advantages, the position of an OMS has been fixed in most of applications, so the camera parameter recovery of an OMS has not been well discussed in the literature. Application fields of OMS would be expanded if absolute position and posture of OMS could be recovered accurately; for example, human navigation, environment virtualization and 3-D model reconstruction. This paper provides an absolute and accurate camera parameter recovery method for a moving OMS using both a few feature landmarks and many natural features.

In a common single camera system, the extrinsic camera parameter reconstruction problem from a single image using n-point feature landmarks of known 3-D and 2-D positions are called the perspective n-point (PnP) problem [7–9].

Generally, this problem can be solved by a least-squares minimization method if six or more feature landmarks can be observed [10]. Although Chen and Chang [11] have extended this problem for generalized projection image sensors, the PnP problem for a multi-camera system has not been well discussed. Additionally, these PnP approaches cannot be successfully used if feature landmarks are partially invisible in the image sequence.

On the other hand, for a single camera system, there is another approach to recovering extrinsic camera parameters from motion of image features [12–14]. They are called the shape-from-motion (SFM). Although these techniques are also attempted for omni-directional camera systems [15–17], they cannot deal with a large number of images because such methods are sensitive to feature tracking errors and these errors must be accumulated. The problem of non-unified center of projection in an OMS has not also been well discussed.

In this paper, a practical, accurate and absolute method for extrinsic camera parameter recovery of an OMS is proposed. In our method, both the PnP and SFM techniques are used tracking both feature landmarks and natural features. The assumptions in our method are that intrinsic camera parameters (including local extrinsic parameters in camera block) are calibrated in advance and are fixed during capturing an image sequence. A small number of feature landmarks need to be visible in some frames of input for minimizing accumulative errors.

This paper is structured as follows. First, the PnP problem for an OMS at a fixed position is discussed in Section 2. Section 3 describes an extrinsic camera parameter recovery method for a moving OMS. Experimental results with real scenes and simulation will show the feasibility and accuracy of the proposed method in Section 4. Finally, Section 5 describes conclusion and future work.

2 PnP problem of OMS

This section describes a method for estimating absolute extrinsic camera parameters of an OMS by solving the PnP problem. In the PnP problem, the extrinsic camera parameters are estimated by using at least six feature landmarks of known 3-D and 2-D positions. Under assumptions of this problem, intrinsic camera parameters (focal length, lens distortion parameters, center of distortion, aspect ratio) and local extrinsic parameters (relative camera positions and postures) of the camera block are known.

In the following sections, first, extrinsic camera parameters of an OMS and projection errors of feature landmarks are defined. The PnP problem is then solved to acquire the extrinsic camera parameters by dealing with multiple cameras and multiple landmarks systematically.

2.1 Extrinsic camera parameters and projection errors

In this section, extrinsic camera parameters of an OMS at a fixed position and projection errors of feature landmarks are defined. Generally, an OMS is composed of multiple cameras and each camera is fixed at a certain place in the

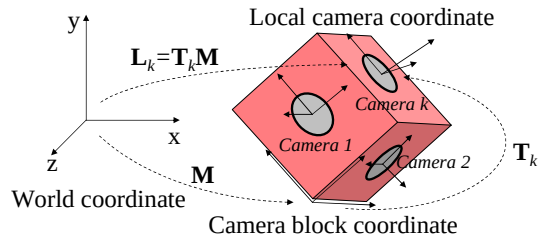


Fig. 1. Coordinate system of an OMS

OMS. As shown in Figure 1, the position and posture of the OMS are determined by the relationship between the camera block coordinate system and the world coordinate system, and those of each camera are determined by the relationship between the local camera coordinate system and the camera block coordinate system. In the following, the world coordinate system is transformed to the camera block coordinate system by a 4×4 matrix $\mathbf{M}$, and the camera block coordinate system is transformed to the local camera coordinate system of the k-th camera by the local extrinsic camera parameter $\mathbf{T}_k$.

First, 6-DOF(degree of freedom) extrinsic camera parameters (posture: r_1, r_2, r_3, position: t_1, t_2, t_3) from the world coordinate to the camera block coordinate are defined as follows.

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1}$$

$$= \begin{pmatrix} \mathbf{R}(r_1, r_2, r_3) & (t_1, t_2, t_3)^T \\ \mathbf{0} & 1 \end{pmatrix}, \tag{2}$$

where $\mathbf{R}$ is a 3×3 rotation matrix. The local extrinsic camera parameter $\mathbf{T}_k$ of the k-th camera is also defined by a 4×4 matrix in the same manner. By using $\mathbf{M}$ and $\mathbf{T}_k$, the extrinsic camera parameter $\mathbf{L}_k$ of the k-th camera in the world coordinate system is expressed as: $\mathbf{L}_k = \mathbf{T}_k\mathbf{M}$.

In the following expressions, for simplicity, we assume that the focal length is 1 and the lens distortion effect is already corrected. The relationship between the 2-D position (u_p, v_p) on the k-th camera image and its 3-D position $\mathbf{S}_p = (x_p, y_p, z_p, 1)^T$ of a feature landmark p in the world-coordinate system can be expressed as:

$$\begin{pmatrix} au_p \\ av_p \\ a \end{pmatrix} = \mathbf{L}_k\mathbf{S}_p = \mathbf{T}_k\mathbf{M}\mathbf{S}_p, \tag{3}$$

where, a is a parameter. A computed position (u_p, v_p) by Eq. (3) and an actually detected position $(\hat{u}_p, \hat{v}_p)$ of the feature landmark p are not always consistent with each other due to quantization and detecting errors. The sum of squared distances between (u_p, v_p) and $(\hat{u}_p, \hat{v}_p)$ for multiple m landmarks has been often

used as an error function in camera parameter estimation for a single camera system [18]. In this paper, the sum of squared errors E is defined as an error function for an OMS as follows.

$$E = \sum_{k=1}^{n} \sum_{p \in \mathbf{F}_k} \{(u_p - \hat{u}_p)^2 + (v_p - \hat{v}_p)^2\}, \tag{4}$$

where $\mathbf{F}_k$ is a set of visible landmarks from the k-th camera, and n denotes the number of camera units in the camera block.

2.2 Solving the PnP problem by minimizing projection errors

This section provides a method for solving the PnP problem of the OMS, which is based on minimizing the error function E defined in Eq. (4) in the previous section. By solving the PnP problem, the extrinsic camera parameter $\mathbf{M}$ of the OMS can be acquired.

The problem of computing $\mathbf{M}$ with 6-DOF by minimizing E is a non-linear problem. To avoid local minimum solutions in a non-linear minimization process, a linear method is first used for computing an initial estimate of $\mathbf{M}$ without the 6-DOF constraint. After that, $\mathbf{M}$ is adjusted to 6-DOF and refined by a gradient method so as to minimize E globally.

Linear estimation of an initial parameter: To solve the PnP problem linearly, multiple n cameras and totally j feature landmarks are used systematically. The local extrinsic camera parameter $\mathbf{L}_k$ of the k-th camera is re-defined using row vectors ($\mathbf{lx}_k$, $\mathbf{ly}_k$, $\mathbf{lz}_k$) as follows.

$$\mathbf{L}_k = \mathbf{T}_k \mathbf{M} = \begin{pmatrix} \mathbf{lx}_k \\ \mathbf{ly}_k \\ \mathbf{lz}_k \end{pmatrix}, \tag{5}$$

Eq. (3) is transformed by Eq. (5) as follows.

$$\begin{aligned} \mathbf{lx}_k \mathbf{S}_p - \hat{u}_p \mathbf{lz}_k \mathbf{S}_p = 0 \ , \\ \mathbf{ly}_k \mathbf{S}_p - \hat{v}_p \mathbf{lz}_k \mathbf{S}_p = 0 \ . \end{aligned} \tag{6}$$

Eq. (6) can be unified about j points of feature landmarks, and transformed by using the parameter vector $\mathbf{m} = (m_{11}, \cdots, m_{14}, m_{21}, \cdots, m_{24}, m_{31}, \cdots, m_{34})^T$ of $\mathbf{M}$ and the local extrinsic camera parameter $\mathbf{T}_k$ as follows.

$$\mathbf{Am} = \mathbf{s}, \tag{7}$$

$$\mathbf{A} = \begin{pmatrix} s1(k_1)\mathbf{S}_1 & s2(k_1)\mathbf{S}_1 & s3(k_1)\mathbf{S}_1 \\ \vdots & \vdots & \vdots \\ s1(k_j)\mathbf{S}_j & s2(k_j)\mathbf{S}_j & s3(k_j)\mathbf{S}_j \\ s5(k_1)\mathbf{S}_1 & s6(k_1)\mathbf{S}_1 & s7(k_1)\mathbf{S}_1 \\ \vdots & \vdots & \vdots \\ s5(k_j)\mathbf{S}_j & s6(k_j)\mathbf{S}_j & s7(k_j)\mathbf{S}_j \end{pmatrix}, \mathbf{s} = \begin{pmatrix} -s4(k_1) \\ \vdots \\ -s4(k_j) \\ -s8(k_1) \\ \vdots \\ -s8(k_j) \end{pmatrix}, \tag{8}$$

$$\begin{pmatrix} s1(k_i) & s2(k_i) & s3(k_i) & s4(k_i) \\ s5(k_i) & s6(k_i) & s7(k_i) & s8(k_i) \end{pmatrix} = \begin{pmatrix} 1 & 0 & \hat{u}_i & 0 \\ 0 & 1 & \hat{v}_i & 0 \end{pmatrix} \mathbf{T}_{k_i} \quad (i = 1, \cdots, j), \tag{9}$$

where k_i is the number associated with the camera from which the feature i is visible.

In Eq. (7), all the parameters except $\mathbf{m}$ are known. If j is six or more, $\mathbf{m}$ can be determined linearly so as to minimize $|\mathbf{Am} - \mathbf{s}|^2$:

$$\mathbf{m} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{s}. \tag{10}$$

Note that, if distances between local cameras are much smaller than 3-D distribution of feature landmarks, computed values in Eq. (10) becomes often unstable. In this case, $\mathbf{m}'$ is defined by $m'_{ij} = m_{ij}/m_{34}$, and by approximating $\mathbf{s}$ as $\mathbf{0}$, stable values will be acquired except the scale parameter of $\mathbf{M}$ by solving $\mathbf{Am}' = \mathbf{0}$.

Camera parameter adjustment to 6-DOF: 12 parameters of $\mathbf{M}$ should be reduced to 3 position parameters (t_1, t_2, t_3) and 3 posture parameters (r_1, r_2, r_3) for Euclidean reconstruction. In this research, the position (t_1, t_2, t_3) of the OMS is simply decided as (m_{14}, m_{24}, m_{34}). The posture parameters (r_1, r_2, r_3) of the OMS can be determined from the rest of 9 parameters of rotation factors $\hat{\mathbf{R}}$ of $\mathbf{M}$ by the singular value decomposition method [19] as follows:

$$\mathbf{R}(r_1, r_2, r_3) = \mathbf{U}diag(1, 1, det(\mathbf{UV}^T))\mathbf{V}^T, \tag{11}$$

where $\mathbf{U}$ and $\mathbf{V}$ are a left and right singular vector matrices of $\hat{\mathbf{R}}$, respectively.

Non-linear minimization of error function: From the initial estimate of the 6 extrinsic camera parameters $(r_1, r_2, r_3, t_1, t_2, t_3)$ acquired by the previous step, The error function E defined in Eq. (4) is minimized using a gradient method by iterating the following expressions until convergence.

$$r_i \leftarrow r_i - l_{r_i}\frac{\delta E}{\delta r_i}, \quad t_i \leftarrow t_i - l_{t_i}\frac{\delta E}{\delta t_i} \ (i = 1, 2, 3) \tag{12}$$

where $(l_{r_1}, l_{r_2}, l_{r_3}, l_{t_1}, l_{t_2}, l_{t_3})$ are scale factors of derivatives, and these values are decided so as to minimize E in each iteration. By using this method, the extrinsic camera parameter $\mathbf{M}$ of the OMS can be determined by a few iterations so as to minimize E globally with 6 DOF, because the initial estimates are expected to be close to the true parameters.

3 SFM of OMS

This section describes a method for estimating extrinsic camera parameters from omni-directional movies acquired by an OMS. The SFM proposed in this paper is not an ego-motion but absolute position estimation method which is based on using both feature landmarks and natural features.

The extrinsic camera parameters of the moving OMS are estimated by solving the PnP problem in each frame of input sequences using both 3-D and 2-D

positions of feature landmarks and natural features. The feature landmarks are used as detonators for the 3-D position estimation of natural features. The 3-D positions of natural features for the PnP are gushed out by tracking them together. In our method, the six or more feature landmarks should be visible at least in the first and the second frames of an input video sequence to acquire the initial estimate of extrinsic camera parameters. Finally, projection errors are globally minimized for all feature landmarks and natural features.

In the following sections, first, error functions in an omni-directional movie are defined. Next the defined errors are minimized to estimate extrinsic camera parameters of the OMS in the input sequence.

3.1 Definition of error function

The sum of projection errors E defined in Eq. (4) is extended for a movie input. The modified sum of projection errors in the f-th frame ($f = 1, \cdots, v$) of the input movie is represented by the following expression.

$$E_f = \sum_{k=1}^{n} \sum_{p \in \mathbf{F}_{kf}} W_p \big\{ (u_{fp} - \hat{u}_{fp})^2 + (v_{fp} - \hat{v}_{fp})^2 \big\}, \tag{13}$$

where W_p is a confidence of feature p, and that is computed as an inverse covariance of projection errors of the feature p [20]. $\mathbf{F}_{kf}$ is a set of feature landmarks and natural features that are visible in the f-th frame image of the k-th camera.

By using E_f, the total error of the input movie is defined as:

$$E_{total} = \sum_{f=1}^{v} A_f E_f, \tag{14}$$

where A_f is a weighting coefficient for each frame f that is set as 1 when the frame contains no specified feature landmarks or else is set to a sufficiently large value when the frame contains specified feature landmarks. In this paper, the error function E_{total} is employed for estimating the extrinsic camera parameters $\mathbf{M}_f (f = 1, \cdots, v)$ and the 3-D positions $\mathbf{S}_p$ of natural features.

On the other hand, the sum of projection errors about the feature p from the fs-th frame to the fe-th frame is also defined as follows:

$$EF_p(fs, fe) = \sum_{f=fs}^{fe} \big\{ (u_{fp} - \hat{u}_{fp})^2 + (v_{fp} - \hat{v}_{fp})^2 \big\}. \tag{15}$$

3.2 Estimation of extrinsic camera parameters from an omni-directional movie

In this section, first, initial parameters of $\mathbf{M}_f$ and $\mathbf{S}_p$ are estimated by tracking both feature landmarks and natural features automatically by using a robust approach. Next, $\mathbf{M}_f$ and $\mathbf{S}_p$ are then refined so as to minimize E_{total} by a non-linear minimization method. The method described in this section is basically an extension of our previous work [20] for the OMS.

Feature tracking and initial estimation: The initial extrinsic parameter of the OMS in each frame is estimated by using the PnP technique described in the Section 2. Our method can compute the initial parameters even if the feature landmarks are invisible in most frames of a long input movie, because the substitutes of feature landmarks with known 3-D and 2-D positions are gushed out by tracking natural features. The feature landmarks are used as detonators for obtaining 3-D positions of natural features. By using a huge number of natural features to solve the PnP problem, accurate and stable estimation can be accomplished. The following paragraphs briefly describe computational steps for the f-th frame.

(a) Feature tracking in each camera: Feature landmarks are tracked automatically by a standard template matching method until a sufficient number of 3-D positions of natural features are estimated. Natural features are automatically detected and tracked by using Harris corner detector [21] for limiting feature position candidates on the images. RANSAC approach [22] is also employed for detecting outliers. In this process, these features are tracked within each camera image.

(b) Extrinsic camera parameter estimation: The 3-D and 2-D positions of feature landmarks and natural features are used for estimating the extrinsic camera parameter $\mathbf{M}_f$. In this step, the error function E_f defined in Eq. (13) is minimized by the method described in the Section 2.2.

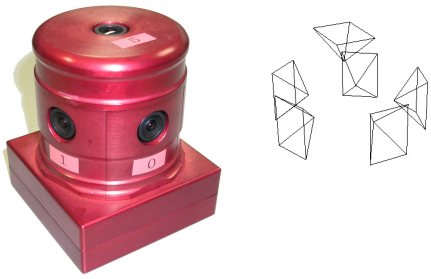
(c) Feature tracking between different camera: The features that become invisible in a certain camera are detected and tracked also in other camera images by using the extrinsic camera parameter $\mathbf{M}_f$ acquired in Step (b). The 3-D position of natural feature that has already been estimated until the previous frame is projected to each camera image by $\mathbf{L}_{kf}(= \mathbf{T}_k\mathbf{M}_f)$, and then the visibility of the 3-D position is checked. If interest points detected by Harris operator exist near by the projected position, the feature is tracked to the nearest interest point.

(d) 3-D position estimation of natural features: The error function EF_p defined in Eq. (15) is used for estimating a 3-D position $\mathbf{S}_p$ of the feature p. For all the natural features tracked in the f-th frame, $EF_p(fs(p), f)$ is minimized and the 3-D position $\mathbf{S}_p$ is refined in every frame, where $fs(p)$ is the first detected frame of the feature p.

(e) Computing confidences of natural features: In this paper, the distribution of tracking errors of the feature is approximated by a Gaussian probability density function. Then, the confidence W_p of the feature p is computed as an inverse covariance of projection errors from the $fs(p)$-th frame to the f-th frame, and refined in every frame [20].

(f) Addition and deletion of natural features: In order to obtain accurate estimates of camera parameters, good features should be selected. In this paper, the set of natural features is automatically updated by checking conditions of features using multiple measures [20].

By iterating the steps above from the first frame to the last frame of the input movie, the initial estimate of the extrinsic camera parameters of the OMS is computed.



(a) appearance (b) viewing volume

Fig. 2. Omni-directional camera system "Ladybug".

Global optimization in video sequence: From the initial parameters of $\mathbf{M}_f (f = 1, \cdots, v)$, the error function E_{total} defined in Eq. (14) is gradually minimized by using derivatives of parameters. This minimization process is almost the same as the method in Section 2.2, except for the 3-D positions of the natural features $\mathbf{S}_p$. In this minimization, the 3-D positions $\mathbf{S}_p = (x_p, y_p, z_p, 1)$ of natural features are also adjustable parameters and refined by using derivatives $(\frac{\delta E_{total}}{\delta x_p}, \frac{\delta E_{total}}{\delta y_p}, \frac{\delta E_{total}}{\delta z_p})$. The feature confidences W_p computed in the iterating process in each frame are also used for this error function E_{total}.

By iterating this minimization for all the input images until convergence, an accurate extrinsic camera parameters and 3-D positions of natural features can be acquired. Local minimum and computational cost problems can be avoided simply by a standard gradient descent method, because the initial parameters are expected to be sufficiently close to the true values.

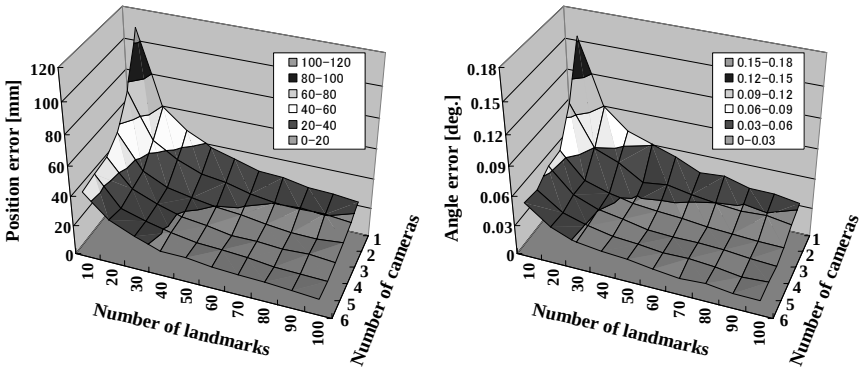
4 Experiments

In this section, two kinds of experimental results are demonstrated. In the first experiment, the accuracy of extrinsic parameters estimated by the method described in Section 2 is evaluated by computer simulation. The second experiment is concerned with 3-D reconstruction test in real environments.

In all the experiments, Ladybug camera system [23] is used as an OMS. As shown in Figure 2, Ladybug has radially located six camera units in the camera block and their positions and postures are fixed. Each camera can acquire 768×1024 resolution images at 15 fps, and the multi-camera system can capture a scene covering more than 75% of full spherical view. The intrinsic parameters of the camera block are estimated as shown in Figure 2(b) by using a laser measurement system and a calibration board [6] in advance. From the result of camera calibration, it is known that displacements of adjacent camera units are 40 ± 2mm in the horizontal direction and 46 ± 4mm in the vertical direction.

4.1 Quantitative evaluation of solving PnP problem in simulation

In this section, the effectiveness of the OMS for the PnP problem is quantitatively evaluated by computer simulation. This simulation is carried out by using a



(a) position error (b) angle error

Fig. 3. Errors in camera block position and angle (simulation).

virtual Ladybug at a fixed position and feature landmarks that are randomly scattered in 50m to 100m range of space from the virtual Ladybug. The feature landmarks are projected to each camera of the virtual Ladybug, and are detected with a Gaussian noise. The Gaussian noise is set so as to have 2.0 pixel standard deviation in each image. After this process, projected positions are quantitized on pixels.

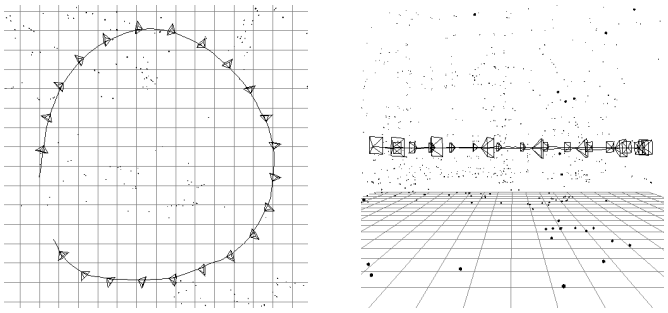
In this situation for solving the PnP problem, both the number of landmarks and the number of cameras of the OMS are changed, and position errors and angle errors by the method described in Section 2 are measured. Figure 3 (a) and (b) show the computed average errors in camera position and angle with a hundred trials. It should be noted that both (a) and (b) exhibit the same behavior: The average error monotonously decreases when the number of landmarks and cameras are increased. Especially about the number of cameras, it is confirmed that the use of the OMS is effective for solving the PnP problem accurately comparing with a single camera system.

4.2 Experiments with real scenes

To demonstrate the validity of the proposed method described in Section 3, extrinsic camera parameters of moving Ladybug are actually reconstructed and evaluated in an indoor and outdoor environment. For both experiments, some of natural features are used as feature landmarks and their 3-D positions are measured by the total station (Leica TCR1105XR). These feature landmarks are specified manually on the first frame image and the last frame image.

Camera parameter recovery for an indoor scene: In this experiment, an indoor scene is captured as an image sequence of 450 frames for each camera by walking in a building as shown in Figure 4. First, the extrinsic camera parameters of Ladybug are reconstructed by the proposed method. On an average, approximately 440 points of natural features are automatically tracked in each

Fig. 4. Sampled frames of input image sequences obtained by six cameras (indoor scene).

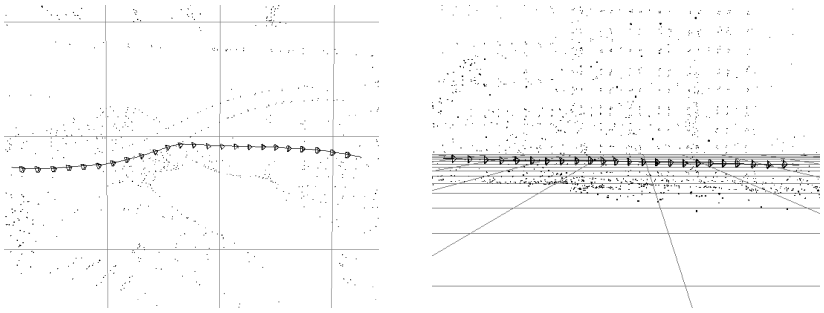


(a) top view (b) side view

Fig. 5. Result of extrinsic camera parameter estimation (indoor scene).

set of frames by six cameras. The squared average of re-projection errors of the features is 2.1 pixels. Figure 5 shows the recovered extrinsic camera parameters of the camera 1 of the Ladybug. The curved line in this figure indicates the camera path, the quadrilateral pyramids indicate the camera postures drawn at every 20 frames. The black point clouds show the estimated 3-D positions of the natural features. The length of the recovered camera path is 29m. As shown in this figure, the camera parameters are recovered very smoothly.

Fig. 6. Sampled frames of input image sequences obtained by six cameras (outdoor scene).



(a) top view (b) side view

Fig. 7. Result of extrinsic camera parameter estimation (outdoor scene).

Camera parameter recovery for an outdoor scene: An outdoor scene is captured by walking in our campus including several buildings as shown in Figure 6. The image sequence for each camera consists of 500 frames. The distances between the camera system and objects in this scene are longer than those in the indoor scene. In this experiment, approximately 530 points of natural features on an average are automatically tracked in each set of frames by six cameras. The squared average of re-projection errors of the features is 1.6 pixels.

Figure 7 shows the recovered extrinsic camera parameters of the camera 1 of the Ladybug. The length of the recovered camera path is also 29m. As same as the indoor environment, the camera parameters are recovered very smoothly.

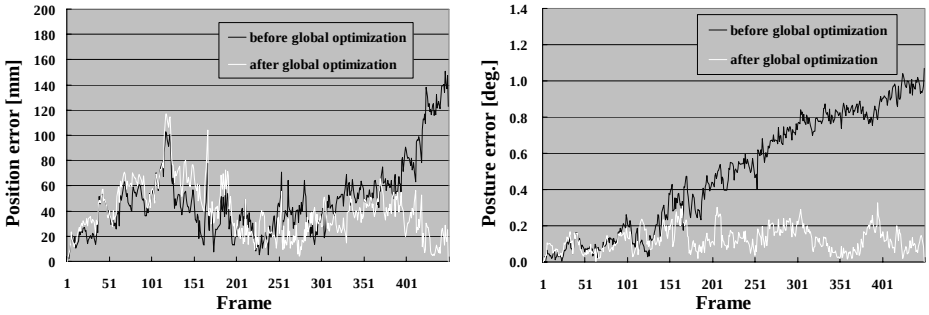
Quantitative evaluation for real scenes: The recovered camera paths and postures are evaluated by comparing with the ground truth. The ground truth is made by solving the PnP problem in every frame. For obtaining the ground truth, features in the input images are manually tracked throughout the input sequences and their 3-D positions are measured by the total station.

Figure 8 denotes position errors and posture errors for the indoor data. The average estimation errors in position and posture of the camera system before global optimization are 50mm and 0.49degree, respectively. After global optimization, they are reduced to 40mm and 0.12degree, respectively. We can confirm that the accumulation of estimation errors is reduced by global optimization. As same as for the indoor sequence, Figure 9 illustrate position errors and posture errors about the outdoor scene. In this sequence, the average estimation error before global optimization is 280mm(position) and 1.10degrees(angle). After the global optimization, they are reduced to 170mm(position) and 0.23degrees(angle), respectively. It is also confirmed that the accumulation of the estimation errors is effectively reduced by global optimization. Note that the average errors of the outdoor scene are larger than those of the indoor scene, because the scale of the outdoor scene is several times larger than the scale of the indoor scene. Although these errors are considered as significantly small comparing with the scale of each scene, more accurate reconstruction could be accomplished by specifying additional feature landmarks, for example, in the middle frame.

5 Conclusion

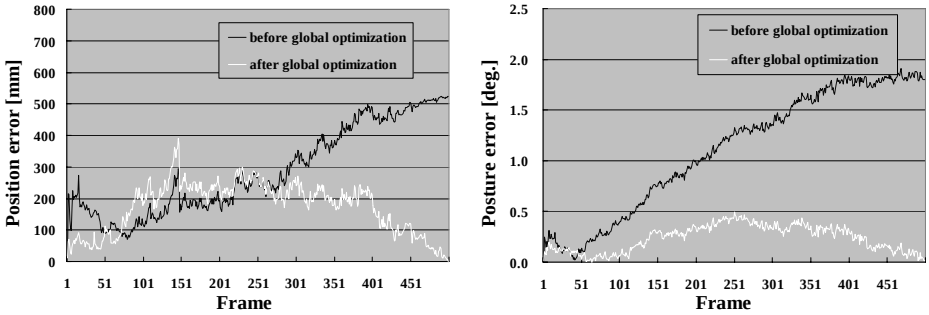
This paper has proposed a method for recovering extrinsic camera parameters of an OMS. In the proposed method, first, the PnP problem for an OMS is solved to recover an extrinsic camera parameter using feature landmarks. Next, extrinsic parameters of the OMS are estimated by using an SFM approach which is based on tracking both feature landmarks and natural features.

In the experiments, the effectiveness of the use of an OMS for PnP problem is quantitatively examined by computer simulation. Additionally, the extrinsic camera parameter recovery with real scenes is successfully demonstrated using multiple long image sequences captured by a real OMS: Ladybug. In future work, the recovered camera parameters will be applied for dense 3-D scene reconstruction of outdoor environments.



(a) position error (b) posture error

Fig. 8. Errors in estimated camera path and posture (indoor scene).



(a) position error (b) posture error

Fig. 9. Errors in estimated camera path and posture (outdoor scene).

References

1. K. Miyamoto: "Fish Eye Lens," Jour. of Optical Society of America, Vol. 54, No. 2, pp. 1060–1061, 1964.
2. K. Yamazawa, Y. Yagi and M. Yachida: "Omnidirectional Imaging with Hyperboloidal Projection," Proc. Int. Conf. on Intelligent Robots and Systems, Vol. 2, pp. 1029–1034, 1993.
3. S. K. Nayar: "Catadioptic Omnidirectional Cameras," Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp. 482–488, 1997.
4. J. Shimamura, H. Takemura, N. Yokoya and K. Yamazawa: "Construction and Presentation of a Virtual Environment Using Panoramic Stereo Images of a Real Scene and Computer Graphics Models," Proc. 15th IAPR Int. Conf. on Pattern Recognition, Vol. IV, pp. 463–467, 2000.
5. H. Tanahashi, K. Yamamoto, C. Wang and Y. Niwa: "Development of a Stereo Omni-directional Imaging System(SOS)," Proc. IEEE Int. Conf. on Industrial Electronics, Control and Instrumentation, pp. 289–294, 2000.
6. S. Ikeda, T. Sato and N. Yokoya: "High-resolution Panoramic Movie Generation from Video Streams Acquired by an Omnidirectional Multi-camera System," Proc.

IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent System, pp. 155–160, 2003.
7. R. Horand, B. Conio and O. Leboullex: "An Analytic Solution for the Perspective 4-Point Problem," Computer Vision, Graphics, and Image Processing, Vol. 47, pp. 33–44, 1989.
8. J. S. C. Yuan: "A General Photogrammetric Method for Determining Object Position and Orientation," IEEE Trans. on Robotics and Automation, Vol. 5, No. 2, pp. 129–142, 1989.
9. R. Krishnan and H. J. Sommer: "Monocular Pose of a Rigid Body Using Point Landmarks," Computer Vision and Image Understanding, Vol. 55, pp. 307–316, 1992.
10. R. Klette, K. Schluns and A. koschan Eds.: Computer Vision: Three-dimensional Data from Image, Springer, 1998.
11. C. S. Chen and W. Y. Chang: "Pose Estimation for Generalized Imaging Device via Solving Non-perspective N Point Problem," Proc. IEEE Int. Conf. on Robotics and Automation, pp. 2931–2937, 2002.
12. P. Beardsley, A. Zisserman and D. Murray: "Sequential Updating of Projective and Affine Structure from Motion," Int. Jour. of Computer Vision, Vol. 23, No. 3, pp. 235–259, 1997.
13. C. Tomasi and T. Kanade: "Shape and Motion from Image Streams under Orthography: A Factorization Method," Int. Jour. of Computer Vision, Vol. 9, No. 2, pp. 137–154, 1992.
14. M. Pollefeys, R. Koch, M. Vergauwen, A. A. Deknuydt and L. J. V. Gool: "Three-dimentional Scene Reconstruction from Images," Proc. SPIE, Vol. 3958, pp. 215–226, 2000.
15. J. Gluckman and S. Nayer: "Ego-motion and Omnidirectional Cameras," Proc. 6th Int. Conf. on Computer Vision, pp. 999–1005, 1998.
16. M. Etoh, T. Aoki and K. Hata: "Estimation of Structure and Motion Parameters for a Roaming Robot that Scans the Space," Proc. 7th Int. Conf. on Computer Vision, Vol. I, pp. 579–584, 1999.
17. C. J. Taylor: "VideoPlus," Proc. IEEE Workshop on Omnidirecitonal Vision, pp. 3–10, 2000.
18. B. Triggs, P. McLauchlan, R. Hartley and A. Fitzgibbon: "Bundle Adjustment a Modern Synthesis," Proc. Int. Workshop on Vision Algorithms, pp. 298–372, 1999.
19. K. Kanatani: Statistical Optimization for Geometric Computation: Theory and Practice, Elsevier Science, 1998.
20. T. Sato, M. Kanbara, N. Yokoya and H. Takemura: "Dense 3-D Reconstruction of an Outdoor Scene by Hundreds-baseline Stereo Using a Hand-held Video Camera," Int. Jour. of Computer Vision, Vol. 47, No. 1-3, pp. 119–129, 2002.
21. C. Harris and M. Stephens: "A Combined Corner and Edge Detector," Proc. Alvey Vision Conf., pp. 147–151, 1988.
22. M.A. Fischler and R.C. Bolles: "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Communications of the ACM, Vol. 24, No. 6, pp. 381–395, 1981.
23. Point Gray Research Inc.: "Ladybug," http://www.ptgrey.com/products/ladybug/index.html.