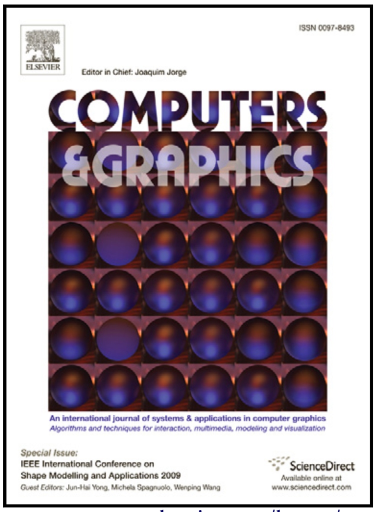
Author's Accepted Manuscript



www.elsevier.com/locate/cag

Camera pose estimation under dynamic intrinsic parameter change for augmented reality

Takafumi Taketomi, Kazuya Okada, Goshiro Yamamoto, Jun Miyazaki, Hirokazu Kato

PII: S0097-8493(14)00064-8
DOI: http://dx.doi.org/10.1016/j.cag.2014.07.003
Reference: CAG2482

To appear in: *Computers & Graphics*

Received date: 7 March 2014
Revised date: 11 July 2014
Accepted date: 11 July 2014

Camera Pose Estimation under Dynamic Intrinsic Parameter Change for Augmented Reality

Takafumi Taketomi[a], Kazuya Okada[b], Goshiro Yamamoto[a], Jun Miyazaki[c], Hirokazu Kato[a]

[a]*Graduate School of Information Science, Nara Institute of Science and Technology*
8916-5 Takayama, Ikoma, Nara 630-0192, Japan
{takafumi-t, goshiro, kato}@is.naist.jp
[b]*Capcom Co., Ltd*
[c]*Department of Computer Science, Tokyo Institute of Technology*
2-12-1 Oookayama, Meguro-ku, Tokyo 152-8552, Japan
miyazaki@cs.titech.ac.jp

Abstract

In this paper, we propose a method for estimating the camera pose for an environment in which the intrinsic camera parameters change dynamically. In video see-through augmented reality (AR) technology, image-based methods for estimating the camera pose are used to superimpose virtual objects onto the real environment. In general, video see-through-based AR cannot change the image magnification that results from a change in the camera's field-of-view because of the difficulty of dealing with changes in the intrinsic camera parameters. To remove this limitation, we propose a novel method for simultaneously estimating the intrinsic and extrinsic camera parameters based on an energy minimization framework. Our method is composed of both online and offline stages. An intrinsic camera parameter change depending on the zoom values is calibrated in the offline stage. Intrinsic and extrinsic camera parameters are then estimated based on the energy minimization framework in the online stage. In our method, two energy terms are added to the conventional marker-based method to estimate the camera parameters: reprojection errors based on the epipolar constraint and the constraint of the continuity of zoom values. By using a novel energy function, our method can accurately estimate intrinsic and extrinsic camera parameters. We confirmed experimentally that the proposed method can achieve accurate camera parameter estimation during camera zooming.

Keywords: Camera Pose Estimation, Augmented Reality, Zoomable Camera, Epipolar Constraint

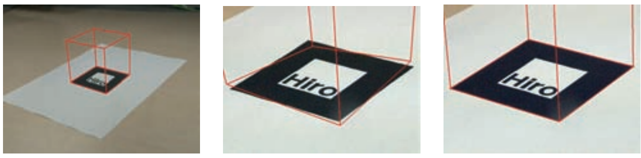
1. Introduction

Augmented reality (AR) is a technique that can integrate the real and virtual worlds. AR enables us to obtain additional information, such as navigation data, guidance, and virtual avatars. Recently, AR applications have been achieved by using a video see-through-based method. In this method, virtual information is overlaid onto a camera image, and the generated AR images are then shown to the user on a display device. In video see-through-based AR applications, geometric registration between the real and virtual worlds is generally required for overlaying the virtual information. Geometric registration for the video see-through-based AR can be achieved by estimating camera parameters.

The methods that are used to estimate camera parameters can be divided into two groups: those for estimating intrinsic camera parameters, including focal length, image center, and lens distortion, and those for estimating extrinsic camera parameters, including camera positions and orientations. In most AR applications, intrinsic camera parameters are calibrated and fixed before the online extrinsic camera parameter estimation process. Many types of methods for estimating camera parameters have been proposed. In these methods, a square marker-based method for estimating extrinsic camera parameters [1] is widely used in various applications, because this method allows the easy construction of a robust AR environment.

Changing the camera's field-of-view, termed "camera zooming," cannot be used in conventional AR applications because intrinsic camera parameters change in the zooming process. Conventional AR applications assume the use of a head-mounted display (HMD) for overlaying virtual information [1, 2]. Camera zooming has not been used for HMDs because zooming gives users an unnatural sensation. This sensation is caused by the difference between the actual head motion and the motion perceived in the displayed images. Thus, the limitation of fixed intrinsic camera parameters in camera parameter estimation is not relevant in conventional AR applications.

In contrast, many types of mobile AR applications for overlaying virtual information that run on smartphones and tablet PCs have been developed recently [3, 4]. In addition, AR technology is often used in the production of TV programs. Although camera zooming in these mobile AR applications or TV programs rarely gives the user an unnatural sensation, these technologies do not allow its use because of the difficulty involved in handling camera zooming in the camera parameter estimation process. Fig. 1(a) shows the results of overlaying a computer-generated (CG) object without camera zooming. Figs. 1(b) and 1(c) show the results of geometric registration



(a) (b) (c)

Figure 1: Example of overlaying a CG object during camera zooming. (a) Without zoom, the intrinsic camera parameters are the same as those used in the calibration process. (b) When the zoom value changes, the CG object is overlaid using the same intrinsic parameters values as in (a). (c) The zoom value was changed. The CG object was overlaid using the proposed method, which considers intrinsic camera parameter changes.

during camera zooming. In the case shown in Fig. 1(b), the registration error increased because inconsistent intrinsic parameters were used to estimate the extrinsic camera parameters. However, in the case shown in Fig. 1(c), accurate geometric registration was achieved by using the proposed method to handle the intrinsic camera parameter change during camera zooming. Removing the limitation caused by fixed intrinsic camera parameters in camera parameter estimation opens possibilities in many AR applications.

To realize simultaneous intrinsic and extrinsic camera parameter estimation during camera zooming, we propose a camera parameter estimation method that uses a pre-calibrated intrinsic camera parameter change and a novel energy function for online camera parameter estimation[1]. In our method, two energy terms are added to the conventional marker-based method for estimating camera parameters: (1) the reprojection errors of tracked natural features and (2) the constraint of the continuity of zoom values. The tracked natural feature points implicitly give a 3D structure of the scene, and the continuity term gives the temporal constraint for the camera parameters. Using the new energy function, our method can accurately and stably estimate intrinsic and extrinsic camera parameters in the online estimation process. Our method requires a pre-calibration process. However, this process needs to be executed only once. Thus, this process does not reduce the usefulness of the proposed method. The remainder of this paper is organized as follows. In Section 2, we discuss related work on image-based camera parameter estimation. The proposed framework is described in Section 3, and a quantitative and qualitative evaluation of its effectiveness is presented in Section 4. Finally, in Section 5, we present the conclusion and future work.

2. Related Work

Many vision-based methods for estimating camera parameters have been proposed in the fields of AR and computer vision. In these methods, camera parameters are estimated by solving the Perspective-n-Point (PnP) problem using 2D-3D corresponding pairs. There are two groups of methods for solving the PnP problem: camera parameter estimation under the conditions of either known or unknown intrinsic camera parameters. Recently, numerous methods have been proposed to solve the PnP problem when the intrinsic camera parameters are known [6, 7, 8, 9, 10, 11]. Most camera parameter estimation methods belong to this category. In AR, 2D-3D corresponding pairs are obtained using a 3D model of the environment or a feature landmark database [12, 13, 14].

Solutions for the PnP problem when the intrinsic camera parameters are not known have also been proposed [15, 16]. These methods can estimate the absolute extrinsic camera parameters and focal length from 2D-3D corresponding pairs. However, in these methods, the accuracy of the estimated camera parameters decreases according to the specific geometric relationship of the points. To solve this problem, Bujnak *et al.* proposed a method for estimating extrinsic camera parameters and focal length that uses a Euclidean rigidity constraint in object space [17]. Furthermore, they improved the computational cost of the method [17] by joining planar and non-planar solvers [18]. The method [18] can be implemented in real time on a desktop computer. However, the accuracy of the estimated camera parameters still decreases in this method when the optical axis is perpendicular to the plane formed by the 3D points. Recently, Kukelova *et al.* proposed the five point-based method [19]. This method can achieve more stable camera parameter estimation than can the method proposed by [18]. However, most marker-based applications use a square marker. In these applications, the camera parameters should be estimated from four 2D-3D corresponding pairs.

Unlike in the PnP problem, to estimate the intrinsic and extrinsic camera parameters corresponding pairs of 2D image coordinates in multiple images are used [20, 21, 22]. These methods are usually used in 3D reconstruction from multiple images, as in the structure-from-motion technique [23]. Although these methods do not need any prior knowledge of the target environment, they cannot estimate absolute extrinsic camera parameters. Sturm proposed a self-calibration method for zoom-lens cameras that uses pre-calibration information [24]. The idea behind this method is similar to that of our proposed method. In this method, intrinsic camera parameters are calibrated and then represented by one parameter. In the online process, the estimation of the intrinsic and extrinsic camera parameters uses this pre-calibration information and is based on the Kruppa equation. However, the solution of the Kruppa equation is not robust to noise, and this method cannot estimate absolute extrinsic camera parameters. These methods are impractical for some AR applications because they require that the user arrange the CG objects and coordinate system manually.

In contrast to the previous methods, the method that we propose accurately and stably estimates the intrinsic and absolute extrinsic camera parameters using an epipolar constraint and a pre-calibrated intrinsic camera parameter change. In our method, a fiducial marker is used to obtain 2D-3D corresponding pairs. Natural feature points that do not have 3D positions are used to stabilize the camera parameter estimation results.

[1]Part of this paper was presented at the International Symposium on Mixed and Augmented Reality, 2013 [5]. In the present paper, we address the auto balancing of each energy term, and we have added a quantitative and qualitative evaluation of the proposed method.

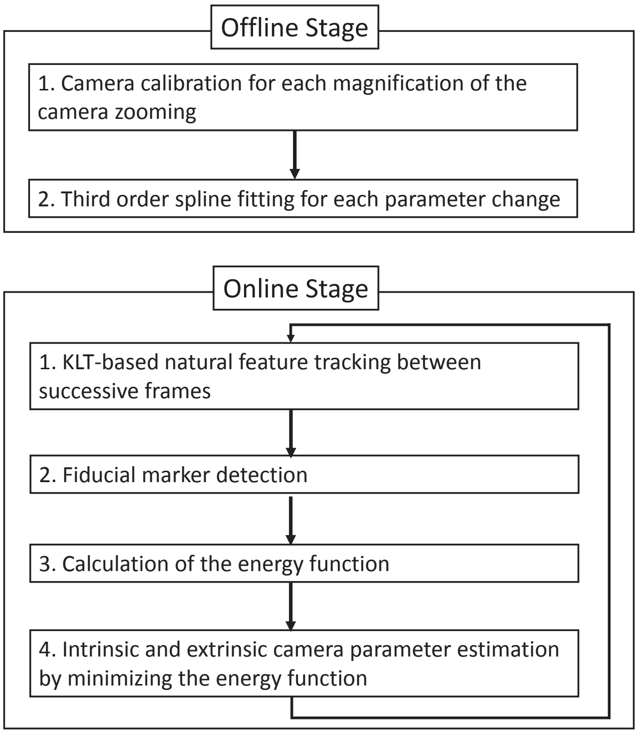


Figure 2: Flow diagram of the proposed method.

Estimated intrinsic camera parameters are constrained by the pre-calibrated intrinsic camera parameter change.

3. Intrinsic and Extrinsic Camera Parameter Estimation for Cameras with Zoom Capabilities

In this section, we describe our method for estimating intrinsic and extrinsic camera parameters in which an energy function is minimized based on the epipolar constraint. Our method is composed of offline camera calibration and online camera parameter estimation, as shown in Fig. 2. Intrinsic camera parameters are modeled using a zoom variable in the calibration process. This model is then used to estimate the camera parameters in the online process. In the online process, several known 3D points and natural features are used to estimate the image magnification that results from the camera zooming and the absolute extrinsic camera parameters. The details of the proposed method are described in the following sections.

3.1. Parameterization of Intrinsic Camera Parameters with Zoom Value

The relationship between the image magnification that results from camera zooming and the intrinsic camera parameters is calibrated in the offline stage. In general, the perspective projection of the pinhole camera model is represented as

$$s\boldsymbol{p} = \boldsymbol{K}\boldsymbol{T}\boldsymbol{P} \tag{1}$$

where $\boldsymbol{P}$ represents the 3D position in the world coordinate system, $\boldsymbol{p}$ is the 2D position in the image coordinate system, and s represents the depth in the camera coordinate system. $\boldsymbol{K}$ and $\boldsymbol{T}$ represent the matrices of the intrinsic and extrinsic camera parameters, respectively. The intrinsic camera parameter matrix has the structure

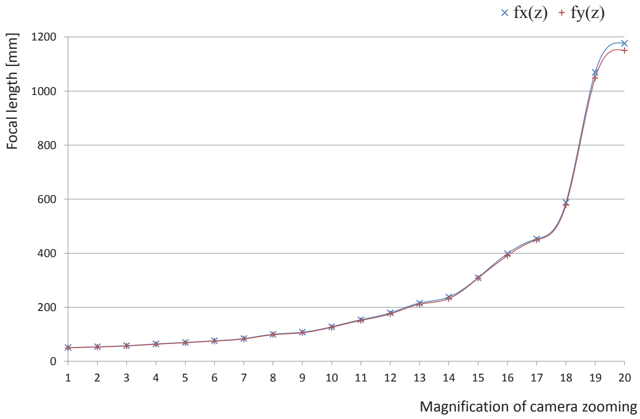


Figure 3: Focal length for each magnification of camera zooming.

$$\boldsymbol{K} = \begin{bmatrix} f_x & 0 & u \\ 0 & f_y & v \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where f_x and f_y are focal lengths, and u and v are the center of projection. In our method, for each image magnification resulting from the camera zooming, these four parameters are measured in an offline camera calibration process. The intrinsic camera parameters are then modeled by the magnification parameter resulting from the camera zooming m. Using this parameterization, we can address the intrinsic camera parameter change using one parameter.

$$\boldsymbol{K}(m) = \begin{bmatrix} f_x(m) & 0 & u(m) \\ 0 & f_y(m) & v(m) \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

In our method, third order spline fitting is used to obtain the model for each parameter change. Figs. 3 and 4 show the calibration results of an intrinsic camera parameter change. The points in each figure indicate the actual parameters obtained by the camera calibration [25]. Each line indicates the third order spline fitting result. We can confirm that the focal lengths are drastically changed at a large image magnification, which results from camera zooming. The accuracy of the geometric registration decreases at this magnification in conventional AR applications.

3.2. Energy Function for Online Camera Parameter Estimation

In the online stage, the Kanade-Lucas-Tomasi (KLT) feature tracker [26] is used and fiducial marker detection is executed. Camera parameters (translation, rotation, and magnification resulting from camera zooming) are then estimated using the information thus obtained. We use the KLT feature tracker

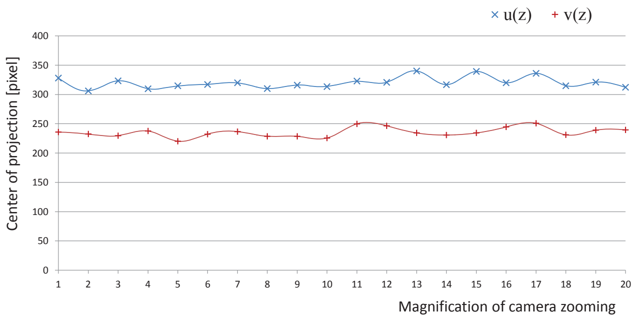


Figure 4: Center of projection for each magnification of camera zooming.

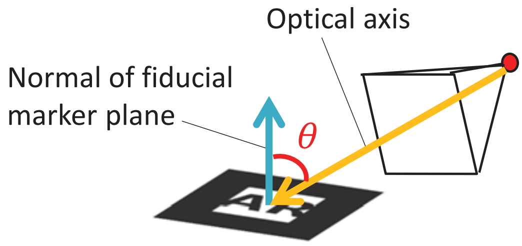


Figure 5: Weight ω_{mk} is calculated according to the angle θ.

because this method can achieve stable feature tracking for image sequences. In addition, its computational cost is relatively low.

To estimate the intrinsic and extrinsic camera parameters in the online stage, we define the new energy function by adding the two energy terms to the conventional marker-based method to estimate camera parameters. The energy function consists of three terms: (1) reprojection errors of the fiducial marker E_{mk}; (2) reprojection errors of tracked natural features based on epipolar constraint E_{ep}; and (3) the constraint of continuity of magnification resulting from camera zooming E_{zoom}. These three terms are automatically balanced using the weights ω_{mk} and ω_{zoom}:

$$E^2 = E_{ep} + \omega_{mk} E_{mk+} + \omega_{zoom} E_{zoom} \tag{4}$$

E_{mk} is used to estimate absolute extrinsic camera parameters, E_{ep} implicitly gives 3D scene structure information, and E_{zoom} gives the temporal constraint for zoom values. The two additional terms improve the accuracy of the estimation of the magnification of the zoom value. In the online process, the camera parameters are estimated by minimizing the energy function E. These terms are described in detail in the following sections.

3.3. E_{mk}: Energy Term based on Fiducial Marker

This energy term is nearly the same as that used in the conventional camera parameter estimation methods. Reprojection errors are calculated from correspondences between the fiducial marker corners in an input image and its reprojected points.

$$E_{mk} = \sum_{i=1}^{4} \left(\boldsymbol{K}(m_j)\boldsymbol{T}_j\boldsymbol{P}_i - \boldsymbol{p'}_i \right)^2 \tag{5}$$

where $\boldsymbol{T}_j$ represents the extrinsic camera parameter matrix composed of camera rotation and translation and $\boldsymbol{P}_i$ and $\boldsymbol{p'}_i$ represent the 3D position of fiducial marker corners and its detected position in the input image, respectively. Unlike in the conventional methods, in the proposed method, the magnification parameter m of the camera zooming exists in the intrinsic camera parameter matrix $\boldsymbol{K}$ in the j-th frame.

It should be noted that the accuracy of marker-based estimation results is unstable when the optical axis of the camera is perpendicular to the fiducial marker plane. This instability is caused by the singularity problem in the optimization process. For this reason, the weight for this energy term ω_{mk} is calculated from the angle θ (Fig. 5) between the optical axis and the fiducial marker plane as

$$\omega_{mk}(\theta) = \frac{4}{\pi^2}\theta^2 + \alpha \tag{6}$$

where α is a minimal weight for E_{mk}.

3.4. E_{ep}: Energy Term based on Epipolar Constraint

E_{ep} is calculated based on the epipolar constraint using natural features tracked between a key and a current frame. In our method, frames that satisfy the following conditions are stored as the key frames.

1. The distance between the current camera position and the camera positions of the previous 10 frames is the maximum.
2. All of the distances between the current camera position and key frame positions are greater than the threshold.

It should be noted that the first frame is stored as the first key frame in the online camera parameter estimation process.

The reprojection errors in term E_{ep} are calculated using natural features tracked between a key frame and the input image. Fig. 6 shows the geometric relationship between two cameras and a corresponding pair of natural features in the input image. In the term E_{ep}, the reprojection error is defined as the distance between an epipolar line l and a detected natural feature position $\boldsymbol{q}_i$ in the input image.

$$E_{ep} = \frac{1}{|\boldsymbol{S}_j|} \sum_{i \in \boldsymbol{S}_j} d_i^2 \tag{7}$$

where $\boldsymbol{S}$ represents a set of tracked natural feature points in the j-th frame, and d_i represents the reprojection error for the natural feature point i. The epipolar line l can be calculated from epipole $\boldsymbol{e}'_i$ and the projected position $\boldsymbol{p}'_i$ of the natural feature position $\boldsymbol{p}_i$ in the key frame. Epipole $\boldsymbol{e}'_i$ and the projected position $\boldsymbol{p}'_i$ are calculated as

$$\boldsymbol{e}'_i = \boldsymbol{K}(m_j)\boldsymbol{T}_j\boldsymbol{P}_{key} \tag{8}$$

$$\boldsymbol{p}'_i = \boldsymbol{K}(m_j)\boldsymbol{T}_j\boldsymbol{P}_i \tag{9}$$

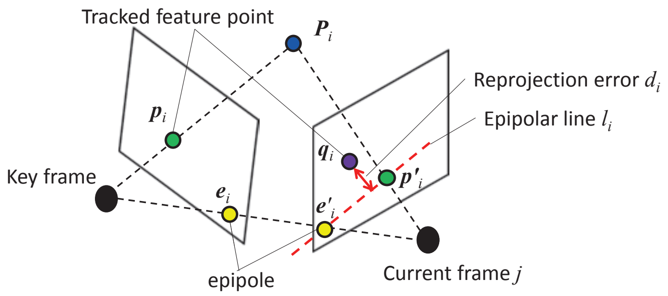


Figure 6: Epipolar constraint between successive frames.

where $\boldsymbol{P}_{key}$ represents the key frame camera position in the world coordinate system. The subscript represents the estimated camera parameters in the key frame. It should be noted that $\boldsymbol{P}_i$ in Eq. (9) is already transformed into the world coordinate system via the matrices $\boldsymbol{K}_{key}\left(m_{key}\right)$ and $\boldsymbol{T}_{key}$. Using this notation, we can represent the estimation error for the two frames based on the epipolar constraint as the reprojection error.

3.5. E_{zoom}: Energy Term based on Continuity of Magnification of Camera Zooming

Our study focuses on the camera parameter estimation for AR. Therefore, the online camera parameter estimation is executed sequentially. In this case, the magnification parameter resulting from camera zooming within successive frames does not change drastically. To use this continuity constraint, we add the following energy term to E_{zoom}:

$$E_{zoom} = \left(m_{j-1} - m_j\right)^2 \tag{10}$$

With this constraint, a discontinuous change in the zoom value is suppressed. It should be noted that the relationship between the zoom values and intrinsic camera parameters is not proportional, as shown in Figs. 3 and 4. These figures show that focal lengths $(f_x(m), f_y(m))$ are drastically changed at a large image magnification as a result of the camera zooming. For this reason, we should control the weight for this term ω_{zoom} adequately. To solve this problem, we employed a weight for ω_{zoom}, which depends on $f_x(m)$ as

$$\omega_{zoom} = \frac{1}{f_x(m_j)} \tag{11}$$

In this term, we only use f_x because the change in f_x is nearly the same as that in f_y.

3.6. Energy Minimization

To estimate the intrinsic and extrinsic camera parameters, the energy function E is minimized using the Levenberg-Marquardt algorithm. The M-estimator is employed in this optimization process to achieve a robust estimation. In this study, we employ the Geman-McClure function ρ.

$$\rho(x) = \frac{x^2/2}{1+x^2} \tag{12}$$

where x represents the residual. The zoom value m_{j-1} estimated in the previous frame and the extrinsic camera parameters estimated using $\boldsymbol{K}\left(m_{j-1}\right)$ are used as initial parameters for the optimization process. In this optimization process, the results of camera parameter estimation sometimes converge at a local minimum. Experimentally, we confirmed that the local minimum problem occurs along the optical axis of the camera. For this reason, to avoid the local minimum problem, the optimization process is executed using three different initial values that are generated by adding an offset β to the initial magnification value of camera zooming. Finally, the lowest energy value of the trial results is chosen, and its estimated camera parameters $\boldsymbol{K}(m_j)$ and $\boldsymbol{T}_j$, are adopted as the final result.

4. Experiment

To demonstrate the effectiveness of the proposed method, we first evaluated the accuracy of the estimated camera parameters in a simulated environment. In this evaluation, the changes in the intrinsic camera parameters during camera zooming were simulated using the measurement results described in Section 4.1. Next, we compared the geometric registration results of our proposed method with those of the state-of-the-art method [18], which can handle camera zooming. The accuracies of the proposed and previous methods were also qualitatively evaluated in the real environment. It should be noted that all input video sequences were started at the non-zoom setting and that the offset for the initial value β in the optimization process was set at 0.1. The value of β was set experientially. In all experiments, we used a desktop PC (CPU: Corei7 2.93 GHz, Memory: 4.00 GB).

4.1. Camera Calibration Results

In this experiment, we used a Sony HDR-AX2000 video camera, which records 640×480 pixel images with an optical zoom (1x-20x) and progressive scan at 30 fps. The lens distortion of this camera is nearly zero ($\kappa_1 = -1.4 \times 10^{-4}$). Thus, we can ignore the lens effect in the following experiments. This video camera was used to generate virtual camera motions in the quantitative evaluation and acquire actual video sequences in the qualitative evaluation. The range of the image magnification resulting from camera zooming is divided into 20 intervals. Then, the intrinsic camera parameters for each zoom value are obtained using Zhang's camera calibration method [25]. Figs. 3 and 4 show the results of the camera calibration. In these figures, the lines indicate the spline fitting results. These results show that the focal length drastically changes when the zoom value is greater than 13. In addition, the center of the projection changes cyclically because the lens rotates during zooming. In the following experiments, we used the spline fitting results of $f_x(z)$, $f_y(z)$, $u(z)$, and $v(z)$.

4.2. Quantitative Evaluation in a Simulated Environment

The accuracy of the estimated intrinsic and extrinsic camera parameters was quantitatively evaluated in a simulated environment. It should be noted that the range of magnification of

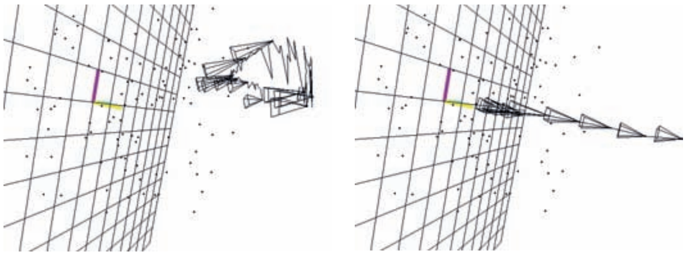
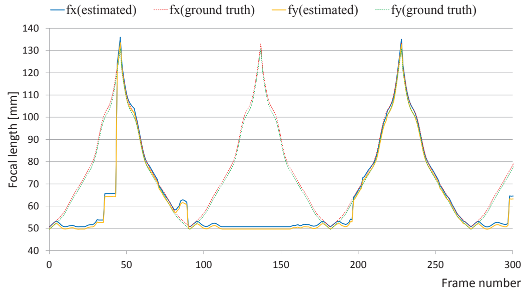
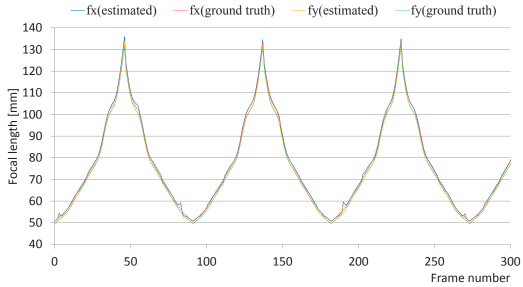


Figure 7: Part of the camera paths and 3D points in the simulated environment. The lefthand figure shows the experimental setup for the free camera motion. The righthand figure shows the experimental setup for the straight camera motion. The fiducial marker center is located at the origin of the world coordinate system, and the marker plane is parallel to the grid plane.



(a)



(b)

Figure 8: Estimation results of focal length for each frame in the case of free camera motion. (a) The estimation results of the previous method [18]. (b) The estimation results of the proposed method.

zooming was reduced from 1x-20x to 1x-10x. This reduction was done because of the difficulty of acquiring ground truth data because the 3D points immediately leave the field of view and the captured images are greatly blurred by the narrow depth of field at large zoom values. In this experiment, the two virtual camera motions used in the simulated environment were acquired using ARToolkit [1] and the video sequences captured in the real environment. In this virtual camera motion acquisition process, intrinsic camera parameters are fixed at the smallest magnification of camera zooming. The differences between these motions are as follows.

- The camera moves freely or straight along the optical axis during camera zooming in the simulated environment.
- The camera travels 2173 mm during free camera motion and 1776 mm during straight camera motion.

In this simulation, 100 3D points were randomly generated in the 3D space (500 mm × 500 mm × 500 mm). Then, the corresponding pairs were obtained by projecting these 3D points into virtual cameras. Additionally, because there was no noise in the projected points, Gaussian noise was added, with the mean equal to zero and a standard deviation of $\sigma = 2.0$. Fig. 7 shows the geometrical relationships between the 3D points and camera motions in the simulated environments.

4.2.1. Free camera motion

In this case, the camera moves freely in the simulated environment, which includes a translation, a rotation, and a zooming. Figs. 8 and 9 show the results of the estimated intrinsic camera parameters (f_x, f_y, u, v) and the ground truth value for each frame. It should be noted that the previous method [18] cannot estimate the centers of the projection. Fig. 9 shows the results of the proposed method only. These results confirm that the proposed method can estimate the focal length more accurately than the previous method. In addition, the proposed method can accurately estimate the center of projection.

Figs. 10 and 11 show the errors for estimated position and rotation. The errors for the camera position are measured by the Euclidean distance between camera centers, while the errors for the camera rotation are measured using the same criteria as those used in [27]. These results confirm that the accuracy of the estimated extrinsic camera parameters is drastically improved by the proposed method. This improvement is considered a result of the accurate estimation of the intrinsic camera parameters. In addition, we can confirm that translation errors are strongly dependent on the zoom factor estimation errors.

Table 1 shows the average errors for each camera parameter. Although the average reprojection error in the previous method is small, the errors for each camera parameter are still large. The results of Figs. 8, 10, 11, and Table 1 confirm that the rotation errors depend on the direction of the optical axis and that the translation errors lie along the optical axis because the resulting reprojection error is small. This is due to the difficulty involved in estimating the parameters using only 2D-3D correspondences. In contrast, the average estimation errors for each camera parameter decrease in the proposed method. We consider that the multiple frame information and the continuity constraint of the camera zooming were responsible for this improvement. However, the processing time of the proposed method is longer than that of the previous method. In the proposed method, the energy minimization process accounts for most of the processing time. To avoid the local minimum problem, in our method, the minimization process is executed for three different initial values. An efficient solver for the energy minimization is needed to allow the proposed method to be adopted in mobile AR applications.

4.2.2. Straight camera motion

In straight camera motion, the camera moves straight along the optical axis during camera zooming. In addition, the optical

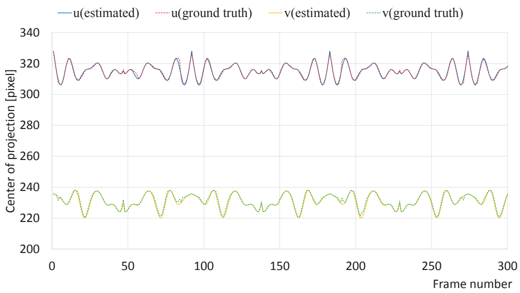


Figure 9: Estimation result of the center of projection for each frame in the case of free camera motion.

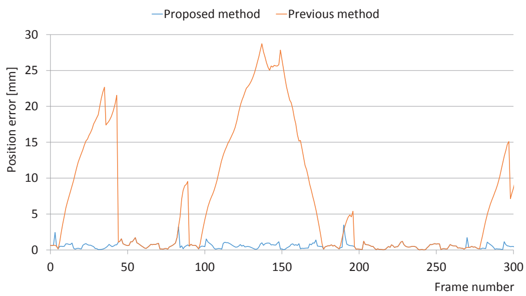


Figure 10: Estimated camera position errors for each frame in the case of free camera motion.

axis is perpendicular to the fiducial marker plane. This condition cannot be easily handled by the previous method [18]. Figs. 12 and 13 show the results of the intrinsic camera parameter estimation. Figs. 14 and 15 show the errors for the estimated position and rotation. Table 2 shows the average errors for each camera parameter. These results show that the proposed method can estimate accurate intrinsic and extrinsic camera parameters under this difficult condition. Conversely, although the reprojection error is small in the previous method, the estimated camera parameters are inaccurate because of the difficulty of estimating camera parameters using only 2D-3D correspondences.

4.2.3. Effect of use of the three initial values

To confirm the effectiveness of the three initial values, we executed the proposed method without offset for the initial value. In this experiment, the camera parameters were estimated using the same input that was used in the free and straight camera motion cases. Fig. 16 shows the results of the focal length estimation in the case of free and straight camera motions obtained using the proposed method without offset for the initial value. In this investigation, we concentrate on the estimation result of the focal length because the accuracies of the other parameters are dependent on the accuracy of the focal length estimation, as shown in Sections 4.2.1 and 4.2.2. The results confirm that the method cannot track the focal length between frames 40 and 50 in the case of straight camera motion. In addition, compared with the results shown in Figs. 8(b) and 12(b), the focal length estimation became unstable. These results show that by using the three initial values in the optimization process, the local minimum problem can be avoided and a more stable camera parameter estimation can be achieved. However, in the proposed method, the computational efficiency is decreased by the three-fold optimization. To reduce the computational cost in the optimization process, efficient initial camera parameter prediction is required.

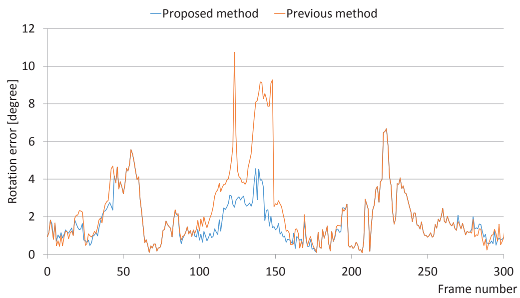


Figure 11: Estimated camera rotation errors for each frame in the case of free camera motion.

4.3. Qualitative Evaluation in the Real Environment

In this experiment, the geometric registration results of the proposed method were compared with those of the previous method [18]. The camera parameter estimation process was executed for two video sequences: one free camera and one straight camera motion sequence. In these sequences, the image magnification resulting from the camera zooming changes dynamically.

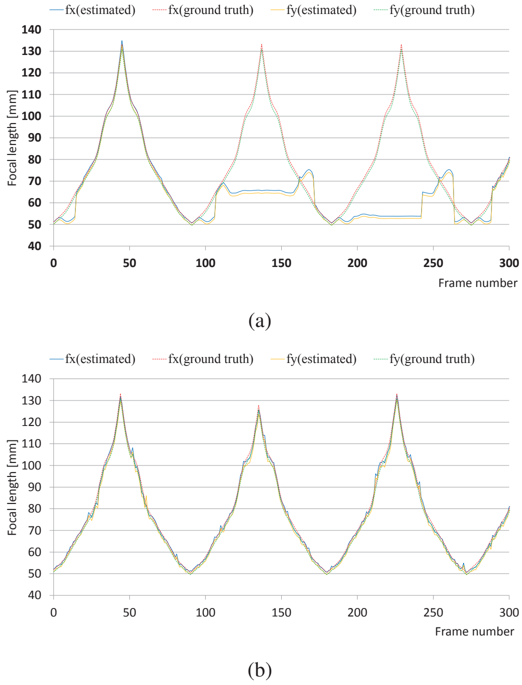
Fig. 17 shows the results of the geometric registration, where a virtual cube is overlaid on a Rubik's cube. We can confirm that the virtual cube is accurately overlaid using the proposed

Table 1: Comparison of accuracy in the case of free camera motion

	Method [18]	Proposed method
Ave. focal length error [mm]	13.66	2.13
Ave. position error [mm]	7.71	1.1
Ave. rotation error [degree]	2.24	1.67
Ave. reprojection error [pixel]	1.33	0.79
Processing time [s]	0.012	0.05

Table 2: Comparison of accuracy in the case of straight camera motion

	Method [18]	Proposed method
Ave. focal length error [mm]	13.08	0.83
Ave. position error [mm]	6.1	0.46
Ave. rotation error [degree]	1.37	1.31
Ave. reprojection error [pixel]	1.36	0.82
Processing time [s]	0.011	0.05



(a)

(b)

Figure 12: Estimation results of focal length for each frame in the case of straight camera motion. (a) The estimation results of the previous method [18]. (b) The estimation results of the proposed method.

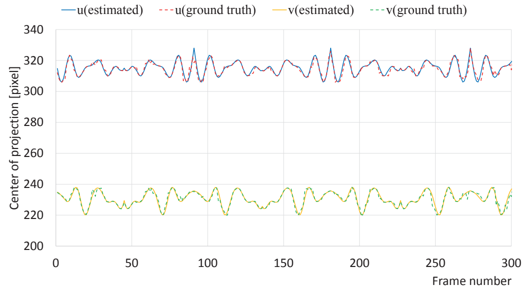


Figure 13: Estimation results of the center of projection for each frame in the case of straight camera motion.

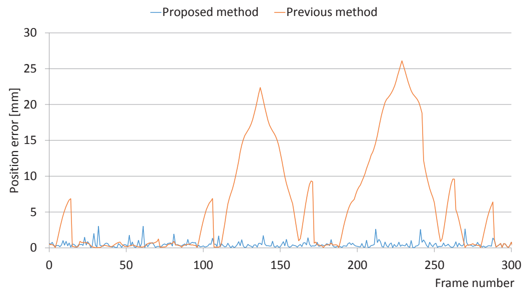


Figure 14: Estimated camera position errors for each frame in the case of straight camera motion.

method. In contrast, the results of the previous method involve geometric inconsistency. More specifically, there is a large geometric inconsistency in the geometric registration results of the previous method for straight camera motion (Fig. 17(b)). These results show that our method can achieve accurate geometric registration using estimated camera parameters even in such a difficult condition.

Fig. 18 shows the results of the estimated camera paths. In this figure, the frustums represent the estimated camera positions and poses. The size of the frustum changes depending on the focal length. This figure confirms that the estimated camera path of the proposed method is smoother than that of the previous method. There is a large jitter in the estimated camera path of the previous method. We confirmed that the proposed method can estimate the camera path more stably than the previous method.

5. Conclusion

In this paper, we proposed a method for estimating a camera pose for environments where the intrinsic camera parameters change dynamically. To estimate intrinsic camera parameters during camera zooming, we developed an energy function based on epipolar geometry. To achieve accurate camera parameter estimation, intrinsic camera parameters at each zoom value are calibrated in advance. Then, the intrinsic camera parameter changes depending on the zoom values are modeled. The effectiveness of the proposed method was demonstrated in simulated and real environments. In the current implementation, our method was applied to a planar scene. However, our method can be applied to non-planar scenes by changing the 2D-3D corresponding pair detection process. This modification allows our method to also be effective in natural feature-based augmented reality applications. If more than four 2D-3D corresponding pairs are used in our method, the term E_{mk} gives a stronger constraint for estimating camera parameters. We did not incorporate lens distortion estimation into the current version of our method because it can be ignored in most consumer cameras. However, when wide angle lenses are used, the lens distortion must be considered; therefore, in future work, lens distortion estimation will be incorporated into the proposed method.

References

[1] Kato H, Billinghurst M. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. Proc Int Workshop on Augmented Reality 1999;:85–94.

[2] Kaufmann H, Dunser A. Summary of usability evaluations of an educational augmented reality application. Proc Int Conf on Virtual Reality 2007;(10):660–9.

[3] Wagner D, Schmalstieg D. First steps towards handheld augmented reality. Proc Int Symp on Wearable Computers 2003;:21–3.

[4] Miyashita T, Meier P, Tachikawa T, Orlic S, Eble T, Scholz V, et al. An augmented reality museum guide. Proc Int Symp on Mixed and Augmented Reality 2008;:103–6.

[5] Taketomi T, Okada K, Yamamoto G, Miyazaki J, Kato H. Geometric registration for zoomable camera using epipolar constraint and pre-calibrated intrinsic camera parameter change. Proc Int Symp Mixed and Augmented Reality 2013;:295–6.

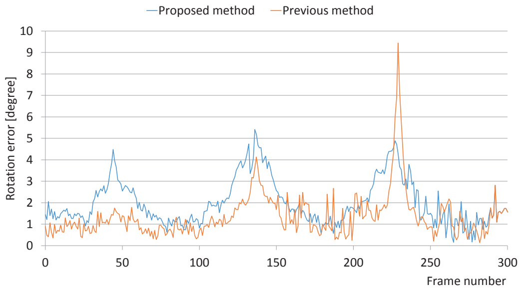
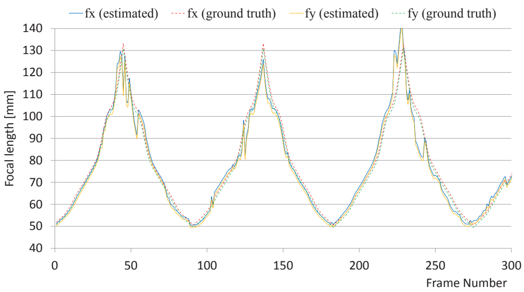
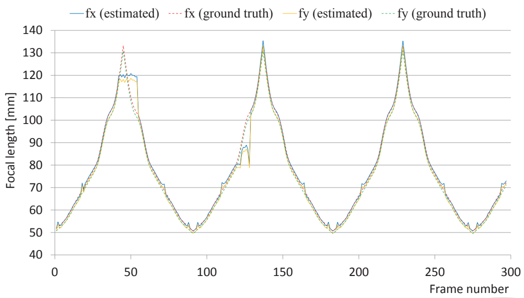


Figure 15: Estimated camera rotation errors for each frame in the case of straight camera motion.



(a)



(b)

Figure 16: Estimation result of the focal length obtained by the proposed method without offset for the initial value. (a) The estimation results for free camera motion. (b) The estimation results for straight camera motion.

[6] Klette R, Schluns K, Koschan A, editors . Computer Vision: Three-dimensional Data from Image. Springer; 1998.

[7] Fischer MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 1981;24(6):381–95.

[8] Quan L, Lan ZD. Linear n-point camera pose determination. IEEE Trans Pattern Analysis and Machine Intelligence 1999;21(8):774–80.

[9] Wu Y, Hu Z. PnP problem revisited. J of Mathematical Imaging and Vision 2006;24(1):131–41.

[10] Lepetit V, Moreno-noguer F, Fua P. EPnP: an accurate $\mathcal{O}(n)$ solution to the PnP problem. Int J of Computer Vision 2009;81(2):155–66.

[11] Hmam H, Kim J. Optimal non-iterative pose estimation via convex relaxation. Int J of Image and Vision Computing 2010;28(11):1515–23.

[12] Drummond T, Cipolla R. Real-time visual tracking of complex structure. IEEE Trans on Pattern Analysis and Machine Intelligence 2002;27(7):932–46.

[13] Lepetit V, Vacchetti L, Thalmann D, Fua P. Stable real-time 3d tracking using online and offline information. IEEE Trans on Pattern Analysis and Machine Intelligence 2004;26(10):1391–402.

[14] Taketomi T, Sato T, Yokoya N. Real-time and accurate extrinsic camera parameter estimation using feature landmark database for augmented reality. Int J of Computers and Graphics 2011;35(4):768–77.

[15] Abidi MA, Chandra T. A new efficient and direct solution for pose estimation using quadrangular targets: Algorithm and evaluation. IEEE Trans Pattern Analysis and Machine Intelligence 1995;17(5):534–8.

[16] Triggs B. Camera pose and calibration from 4 or 5 known 3D points. Proc Int Conf on Computer Vision 1999;:278–84.

[17] Bujnak M, Kukelova Z, Pajdla T. A general solution to the P4P problem for camera with unknown focal length. Proc IEEE Conf on Computer Vision and Pattern Recognition 2008;:1–8.

[18] Bujnak M, Kukelova Z, Pajdla T. New efficient solution to the absolute pose problem for camera with unknown focal length and radial distortion. Proc Asian Conf on Computer Vision 2010;:11–24.

[19] Kukelova Z, Bujnak M, Pajdla T. Real-time solution to the absolute pose problem with unknown radial distortion and focal length. Proc Int Conf on Computer Vision 2013;:2816–23.

[20] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision. Second ed.; Cambridge University Press; 2004.

[21] Stewenius H, Nister D, Kahl F, Schaffalitzky F. A minimal solution for relative pose with unknown focal length. Proc IEEE Conf on Computer Vision and Pattern Recognition 2005;:789–94.

[22] Li H. A simple solution to the six-point two-view focal-length problem. Proc European Conf on Computer Vision 2006;4:200–13.

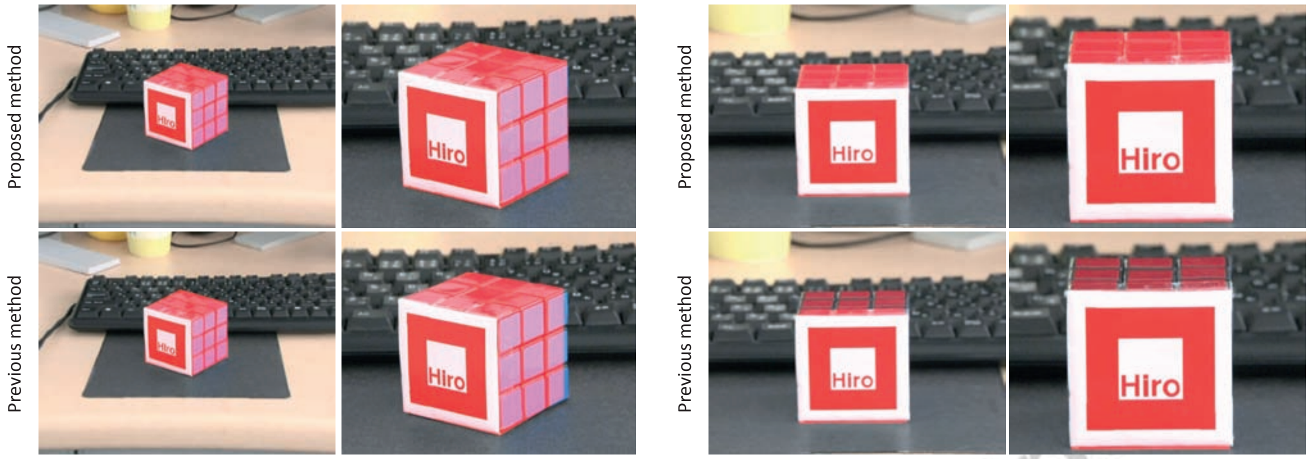
[23] Snavely N, Seitz SM, Szeliski R. Photo tourism: Exploring photo collections in 3D. ACM Trans on GRAPHICS 2006;:835–46.

[24] Sturm P. Self-calibration of a moving zoom-lens camera by pre-calibration. Int J of Image and Vision Computing 1997;15:583–9.

[25] Zhang Z. A flexible new technique for camera calibration. IEEE Trans on Pattern Analysis and Machine Intelligence 2000;22(11):1330–4.

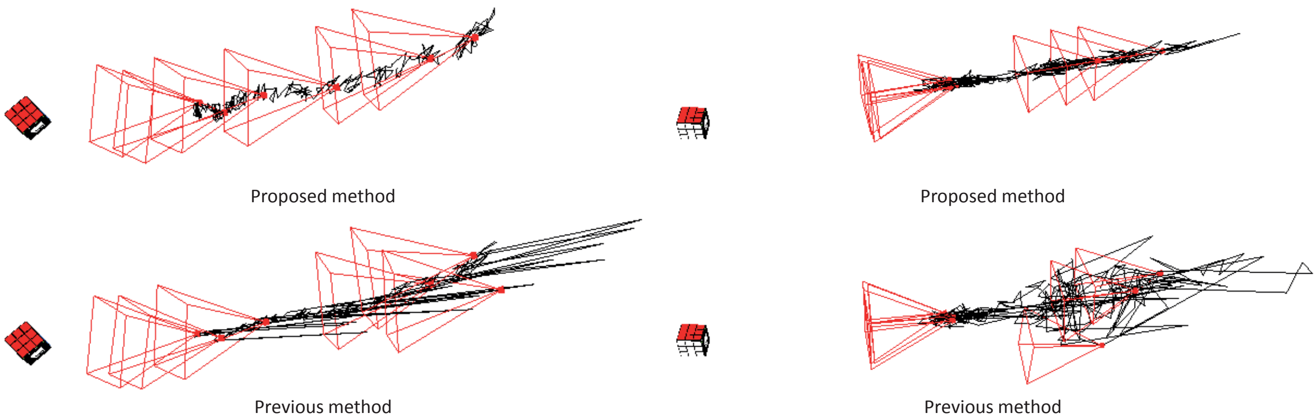
[26] Shi J, Tomasi C. Good features to track. Proc IEEE Conf on Computer Vision and Pattern Recognition 1994;:593–600.

[27] Petit A, Caron G, Uchiyama H, Marchand E. Evaluation of model based tracking with trakmark dataset. Proc Int Workshop on AR/MR Registration, Tracking and Benchmarking 2011;.



(a) (b)

Figure 17: A virtual cube is overlaid on the Rubik's cube in each frame. (a) The geometric registration result for free camera motion. (b) The geometric registration result for straight camera motion. In the top row in each figure, the results of the proposed method are presented for different magnifications as a result of the camera zooming. In the bottom row, the results obtained using the previous method are shown.



(a) (b)

Figure 18: Estimated camera paths. (a) The estimation result for free camera motion. (b) The estimation result for straight camera motion.