FOCAL LENGTH CHANGE COMPENSATION FOR MONOCULAR SLAM

Takafumi Taketomi

Nara Institute of Science and Technology, Japan

Janne Heikkilä

University of Oulu, Finland

ABSTRACT

In this paper, we propose a method for handling focal length changes in the SLAM algorithm. Our method is designed as a pre-processing step to first estimate the change of the camera focal length, and then compensate for the zooming effects before running the actual SLAM algorithm. By using our method, camera zooming can be used in the existing SLAM algorithms with minor modifications. In the experiments, the effectiveness of the proposed method was quantitatively evaluated. The results indicate that the method can successfully deal with abrupt changes of the camera focal length.

Index Terms— SLAM, Camera Zoom, Augmented Reality

1. INTRODUCTION

In augmented reality (AR), camera pose estimation is necessary for achieving geometric registration between the real and virtual worlds. Many kinds of camera pose estimation methods have been proposed in the AR and computer vision research fields. Especially, SLAM-based camera pose estimation is an active research topic.

The SLAM-based camera pose estimation method estimates camera pose and 3D structure of the target environment simultaneously. The SLAM algorithms are composed of a tracking process and a mapping process. Natural features in input images are tracked in successive frames, and 3D positions of natural features are estimated in the mapping process. In general, intrinsic camera parameters are calibrated in advance and these parameters are fixed in the SLAM-based camera pose estimation process. This assumption means that the SLAM algorithms do not allow to use a camera zooming, because that would change the camera focal length.

In the computer vision research, many types of camera parameter estimation methods have been proposed. These methods can be divided into two groups: camera parameter estimation for known and unknown 3D references. The latter is also often referred to as auto-calibration or self-calibration.

Camera parameter estimation from 2D-3D correspondences is known as a Perspective-n-Point (PnP) problem. Many methods for solving the PnP problem have been proposed when the intrinsic camera parameters are unknown [1, 2, 3, 4, 5]. These methods can estimate the focal length

and extrinsic camera parameters, but they cannot be used in the unknown environment because all of these methods need several 3D reference points.

Camera parameter estimation methods from 2D-2D correspondences have also been proposed [6, 7, 8]. They are usually used in offline 3D reconstruction, such as the structure-from-motion technique [9]. Although camera parameter estimation from 2D-2D correspondences is possible in unknown environments, these methods are not suitable for SLAM algorithms. For example, the method [6] needs projective reconstruction in advance, and the methods [7, 8] consider two view constraint only. On the other hand, pre-calibration based methods have been proposed [10, 11]. These methods can estimate the focal length and the extrinsic camera parameters accurately using the dependency of each intrinsic camera parameters. In order to make a lookup table of the intrinsic camera parameter dependency, intrinsic camera parameters at each magnification of camera zooming are calibrated in advance. Although the pre-calibration information gives strong constraint in an online camera parameter estimation process, the pre-calibration process decreases the usability of the application.

In this research, we focus on SLAM-based camera pose estimation, and we propose a method for handling the focal length change caused by camera zooming. The proposed method is designed as a preprocessing step of the SLAM algorithm. The camera zooming effect in the current image is compensated for by using the estimated focal length change, as shown in Fig. 1. By using the proposed preprocessing method, the existing SLAM algorithms can handle camera zooming.

2. REMOVING THE CAMERA ZOOMING EFFECT

The method is composed of four parts, as shown in Fig. 2. In our method, we assume that the principal point is located at the center of the image, aspect ratio is unity, skew is zero, and lens distortion can be ignored. In addition, we assume fixed intrinsic camera parameters in the initialization process of the SLAM algorithm. These assumptions are reasonable for the current consumer camera devices and the SLAM algorithm.
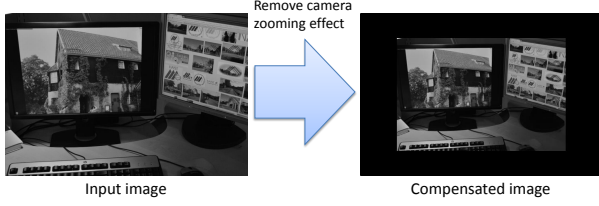
Fig. 1. Image compensation for removing the camera zooming effect. The left image is an input image. The right image is an compensated image by using the estimated focal length change.
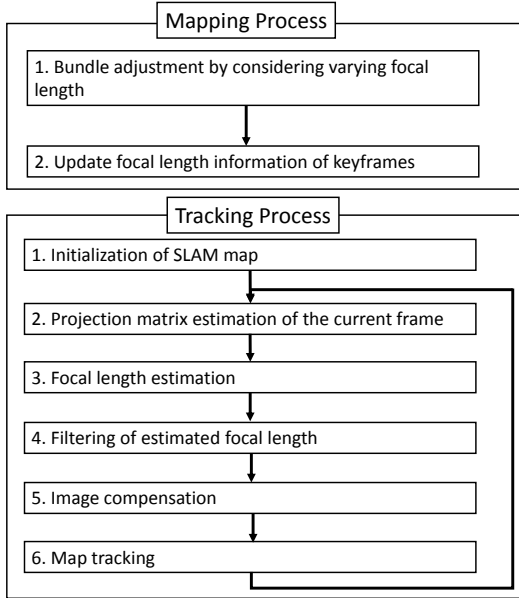
Fig. 2. Flow diagram of the proposed method.

2.1. Focal Length Change Estimation

The focal length change estimation process is based on the method described in [12]. In this method, focal lengths of each image are estimated from projection matrices of the cameras. Basically, this method has been designed for offline metric reconstruction because projective reconstruction is needed before focal length estimation. We extended this method to achieve sequential focal length estimation. In our approach, the projection matrix of the current frame is estimated using tracked natural features. The focal length change is determined based on the estimated projection matrix and the projection matrices of the keyframes.

Projection Matrix Estimation: In order to estimate the projection matrix of the current frame, natural features used for estimating camera parameters of the previous frame are tracked by using the Lucas-Kanade tracker [13]. By using these tracked features, the projection matrix $\boldsymbol{M}$ of the current frame can be estimated by minimizing the following cost function [14].

$$E_p = \sum_{i \in \boldsymbol{S}} ||\boldsymbol{x}_i - proj(\boldsymbol{X}_i)||^2 \tag{1}$$

where $\boldsymbol{S}$ represents a set of tracked natural feature points in the current frame, and $\boldsymbol{x}_i$ represents the image coordinates of the tracked natural feature i, and $proj()$ is a function for projecting the 3D point $\boldsymbol{X}_i$ onto the image using $\boldsymbol{M}$. The initial estimate of the projection matrix $\boldsymbol{M}$ is obtained with a linear algorithm and then the cost function is minimized by using the Levenberg-Marquardt algorithm.

Focal Length Estimation: The focal length of the current frame is estimated from the projection matrices of the current frame and the keyframes. To estimate the focal length, at least three view points are needed [12]. In the map initialization process, two keyframes are used for estimating initial 3D points by a stereo measurement [15]. Because we already have two keyframes after the initialization process, the focal length estimation can be done in real-time during the tracking process.

First, the keyframes that have been used for determining 3D positions of tracked natural features are selected from the map. In addition, the first keyframe which is used for initialization is always selected to provide the reference focal length. The relationship between intrinsic camera parameters and projection matrices of the selected keyframes and the current frame can be described as follows:

$$\boldsymbol{K}_i\boldsymbol{K}_i^T = \boldsymbol{M}_i\Omega^*\boldsymbol{M}_i^T \tag{2}$$

where Ω^* is the absolute quadric that has the 4×4 matrix structure. Intrinsic camera parameter matrices $\boldsymbol{K}_i$ and the absolute quadric Ω^* can be calculated using the rank 3 constraint [12].

Magnification of camera zooming can be estimated from the focal length ratio $f_{1,t}$ between focal lengths of the first keyframe f_1 and the current frame f_t as follows:

$$f_{1,t} = f_1/f_t \tag{3}$$

It should be noted that the focal length ratio $f_{1,t}$ can be regarded as the absolute focal length value because there is a scale ambiguity in SLAM-based reconstruction. If the initial focal length is assumed to be 1, the focal length ratio becomes the value of the focal length in the successive frames.

2.2. Robust Filtering

The focal length estimation process is sensitive to estimation errors of the projection matrices. In order to achieve stable focal length ratio, we employ two filtering processes: median filtering for robust estimation and temporal filtering for smooth estimation.

Median Filtering for Robust Focal Length Estimation: In order to achieve stable estimate, we employ the median filter for estimated focal length ratios obtained by Sec. 2.1. In the
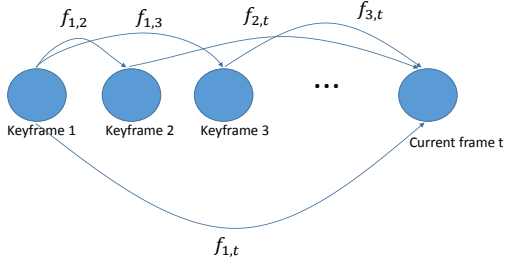
Fig. 3. Focal length ratio estimation by median filtering.

focal length estimation process, the focal length ratio between the first keyframe and the current frame $f_{1,t}$ is estimated, and the focal length ratios between other keyframes and the current frame $f_{2,t}, f_{3,t}, \ldots, f_{n,t}$ are also estimated as shown in Fig. 3 (n represents the number of selected keyframes). In addition, focal length ratios between the first keyframe and the other keyframes $f_{1,2}, f_{1,3}, \ldots, f_{1,n}$ have already been estimated before the focal length estimation process of the current frame.

By using these values, we can obtain candidates of the focal length ratio between the first keyframe and the current frame as follows:

$$f_{1,t}, f_{1,2}f_{2,t}, f_{1,3}f_{3,t}, \ldots, f_{1,n}f_{n,t} \qquad (4)$$

The median value of these candidates is selected as the focal length ratio between the first keyframe and the current frame $f_{1,t}$.

Temporal Filtering for Smooth Estimation: After median filtering the focal length ratio still contains some noise that would cause annoying jitter between frames. In order to reduce the effect of the noise we employ temporal filtering for smoothing the estimate. The estimated focal length ratio is filtered by the following equation.

$$\hat{f}_{1,t} = \alpha f_{1,t} + (1 - \alpha)\hat{f}_{1,t-1} \qquad (5)$$

where $\hat{f}$ represents the filtered focal length ratio and α represents a coefficient for smoothing. The actual focal length ratio can change in successive frames. In order to tolerate smooth changes, we define the following criteria.

- $|f_{1,t} - \hat{f}_{1,t-1}| < \varepsilon_1$: Estimated focal length ratio of the current frame should be similar to that of the filtered previous value.

- $|f_{1,t} - f_{1,t-1}| < \varepsilon_2$: Similar focal length ratios are estimated in the current and previous frames.

- $|f'_{1,t} - f'_{1,t-1}| < \varepsilon_3$: Gradients of estimated focal lengths are similar. Gradients are calculated by $f'_{1,t} = f_{1,t} - f_{1,t-1}, f'_{1,t-1} = f_{1,t-1} - f_{1,t-2}$.

The second and third conditions are for detecting the focal length change. If the estimated focal length ratio f_t satisfies one or more conditions, $f_{1,t}$ is accepted and used in the filtering process (Eq.(5)). If all conditions are false, the filtered focal length ratio of the previous frame is used as an input to the filtering process $f_{1,t} = \hat{f}_{1,t-1}$. In addition, sometimes the focal length ratio cannot be acquired by the focal length estimation method described in Sec.2.1. This happens when the solution for f_i^2 in Eq.(2) has a negative value. The filtered focal length ratio of the previous frame is also used in Eq.(5) when $f_i^2 < 0$. Finally, the input image is scaled using the filtered focal length ratio $\hat{f}_{1,t}$.

2.3. Bundle Adjustment

In bundle adjustment which is a part of the mapping process shown in Fig. 2, changes of the focal length should be also compensated for. In the proposed method, we modify the cost function for dealing with the scale factor which means the error of focal length ratio estimation in the online process.

$$E = \sum_{i \in \boldsymbol{F}} \sum_{j \in \boldsymbol{P}} \left|\left|\boldsymbol{x}_{ij} - proj_i(\boldsymbol{X}_j)\right|\right|^2 \qquad (6)$$

where $\boldsymbol{F}$ and $\boldsymbol{P}$ represent a set of keyframes and a set of reconstructed 3D points respectively. $proj_i()$ represents projection of 3D points $\boldsymbol{X}_j$ onto the keyframe i. 3D points are projected using extrinsic and intrinsic camera parameters.

$$\bar{\boldsymbol{x}}_{ij} \propto s_i [\boldsymbol{R}_i | \boldsymbol{t}_i] \boldsymbol{X}_j \qquad (7)$$

where $\boldsymbol{R}_i$ and $\boldsymbol{t}_i$ represent rotation and translation components respectively, and s_i represents the scale factor for the keyframe i. $\bar{\boldsymbol{x}}_{ij}$ represents the projected position of $\boldsymbol{X}_j$ in the image coordinate system. Solutions for $\boldsymbol{R}_i$, $\boldsymbol{t}_i$, s_i, and $\boldsymbol{X}_j$ are calculated by minimizing the cost function E using non-linear optimization method such as the Levenberg-Marquardt algorithm. After the optimization process, the focal length ratio of each keyframe is updated.

$$f_{i,new} = f_{i,old}/s_i \qquad (8)$$

3. EXPERIMENT

To demonstrate the effectiveness of the proposed method, the accuracy of focal length estimation was quantitatively evaluated. In the experiment, we used PTAM [15] as an existing SLAM algorithm. In all experiments, the hardware included a desktop PC (CPU: Corei5-3570 3.4 GHz, Memory: 8.00 GB) and a Sony NEX-VG900 video camera, which records 640×480 pixel images with an optical zoom lens (Sony SEL1018, $f = 10mm - 18mm$).

In this experiment, the accuracy of estimated focal length ratio is evaluated with two sequences: non-zoom sequence and zoomed sequence. In the both experiments, first 300 frames are used for initialization, and the focal length is set at a fixed value 1.0.
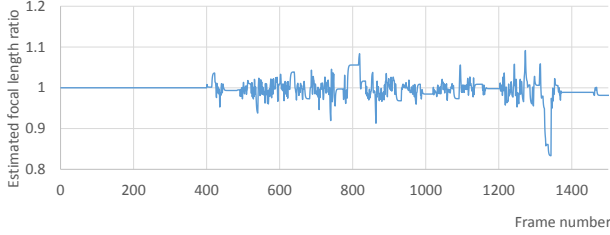
Fig. 4. The estimation result of focal length ratio in non-zoom sequence.
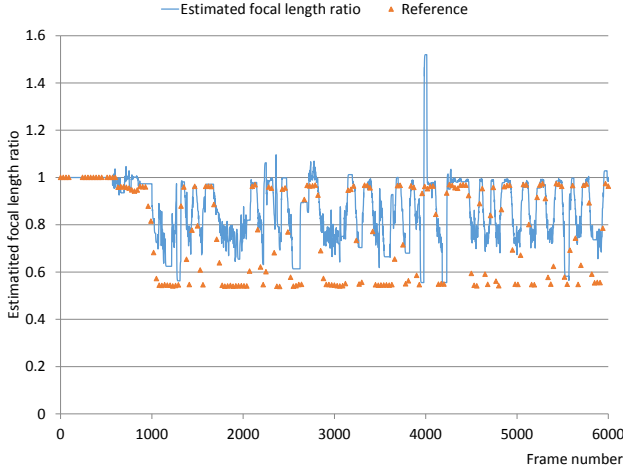

Fig. 5. The estimation result of focal length ratio in zoomed sequence.

Non-zoomed Sequence: In this case, the camera moves freely in the real environment, which includes translation and rotation. A maximum distance between the camera and the target scene was about 2 meters.

Fig. 4 shows the result of focal length estimation. In this figure, estimated focal length ratios should lie at 1. An average error for focal length estimation was 0.012 and its standard deviation was 0.019. This result confirms that the focal length of the input image was accurately estimated. It also indicates that the proposed method does not have much effect on the accuracy of the conventional SLAM algorithm.

Zoomed Sequence: In this case, the camera moves freely in the real environment, which includes translation, rotation, and camera zooming. In order to evaluate the accuracy of focal length estimation, reference focal length values for each image were obtained by an offline reconstruction method [16, 17]. The reference values were obtained at every 30th frames. Figs. 5 and 6 show the result of focal length estimation and its estimation errors in each frame respectively. In Fig. 5, triangle points represent reference focal length ratio obtained from offline reconstruction. An average error for focal length estimation was 0.113 and its standard deviation was 0.109. The result confirms that the proposed method can estimate the focal length change with reasonable accuracy. However, estimated focal length ratio involves a small delay. This delay
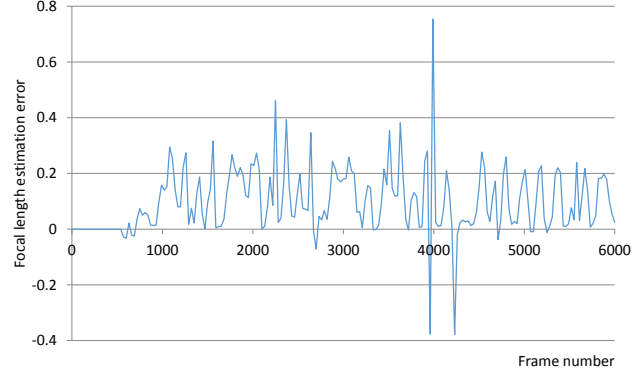
Fig. 6. Focal length estimation error in each frame.

is caused by the temporal filtering process.

In addition, we can observe a large spike around the 4000th frame. At this time, the camera moved along the optical axis with simultaneous zooming. In general, zooming and translation along the optical axis cause an ambiguity which is difficult to handle especially if the scene structure is relatively flat. For SLAM this is probably a rare case, and it could be avoided by adding more heuristics to the algorithm.

The execution time for our preprocessing algorithm is shown in Table 1. A half of the processing time for estimating the projection matrix was used by the Lucas-Kanade tracker (5.51 ms). The result confirms that the proposed method still can work in realtime.

4. CONCLUSION

In this paper, we proposed a focal length change compensation method for dealing with camera zooming in SLAM algorithms. The main benefit of this method is that the camera zooming effect in the input image can be compensated before the tracking process in SLAM algorithm which enables use of existing SLAM algorithms together with our method. In order to estimate the focal length change, we developed an online focal length estimation framework. In this framework, the estimated focal length is filtered in two stages to achieve more stable result. The effectiveness of the proposed method was demonstrated in the experiments.

Table 1. Average computational time for each process.

Process	time (ms)
Projection matrix estimation	11.78
Focal length estimation	0.08
Robust filtering	0.51
Image compensation	0.27
Map tracking	13.58
Total	26.22

5. REFERENCES

[1] M A. Abidi and T. Chandra, "A new efficient and direct solution for pose estimation using quadrangular targets: Algorithm and evaluation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 534–538, 1995.

[2] B. Triggs, "Camera pose and calibration from 4 or 5 known 3D points," *Proc. Int. Conf. on Computer Vision*, pp. 278–284, 1999.

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

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

[5] Z. Kukelova, M. Bujnak, and T. Pajdla, "Real-time solution to the absolute pose problem with unknown radial distortion and focal length," *Proc. Int. Conf. on Computer Vision*, pp. 2816–2823, 2013.

[6] Marc Pollefeys, Reinhard Koch, and Luc Van Gool, "Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters," *Int. J. of Computer Vision*, pp. 7–25, 1999.

[7] H. Stewenius, D. Nister, F. Kahl, and F. Schaffalitzky, "A minimal solution for relative pose with unknown focal length," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 789–794, 2005.

[8] H. Li, "A simple solution to the six-point two-view focal-length problem," *Proc. European Conf. on Computer Vision*, vol. 4, pp. 200–213, 2006.

[9] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3D," *ACM Trans. on GRAPHICS*, pp. 835–846, 2006.

[10] P. Sturm, "Self-calibration of a moving zoom-lens camera by pre-calibration," *Int. J. of Image and Vision Computing*, vol. 15, pp. 583–589, 1997.

[11] T. Taketomi, K. Okada, G. Yamamoto, J. Miyazaki, and H. Kato, "Camera pose estimation under dynamic intrinsic parameter change for augmented reality," *Computers and Graphics*, vol. 44, pp. 11–19, 2014.

[12] Marc Pollefeys, Reinhard Koch, and Luc Van Gool, "Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters," *Int. J. of Computer Vision*, pp. 7–25, 1999.

[13] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proc. of Int. Joint Conf. on Artificial Intelligence*, pp. 674–679, 1981.

[14] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, second edition, 2004.

[15] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 225–234, 2007.

[16] Changchang Wu, "Towards linear-time incremental structure from motion," *Proc. Int. Conf. on 3D Vision*, pp. 127–134, 2013.

[17] Changchang Wu, S. Agarwal, B. Curless, and S.M. Seitz, "Multicore bundle adjustment," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3057–3064, 2011.