Extrinsic Camera Parameter Estimation Using Video Images and GPS Considering GPS Positioning Accuracy

Hideyuki Kume, Takafumi Taketomi, Tomokazu Sato and Naokazu Yokoya
Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0192, Japan
{*hideyuki-k, takafumi-t, tomoka-s, yokoya*}*@is.naist.jp*

***Abstract*—This paper proposes a method for estimating extrinsic camera parameters using video images and position data acquired by GPS. In conventional methods, the accuracy of the estimated camera position largely depends on the accuracy of GPS positioning data because they assume that GPS position error is very small or normally distributed. However, the actual error of GPS positioning easily grows to the 10m level and the distribution of these errors is changed depending on satellite positions and conditions of the environment. In order to achieve more accurate camera positioning in outdoor environments, in this study, we have employed a simple assumption that true GPS position exists within a certain range from the observed GPS position and the size of the range depends on the GPS positioning accuracy. Concretely, the proposed method estimates camera parameters by minimizing an energy function that is defined by using the reprojection error and the penalty term for GPS positioning.**

***Keywords*-extrinsic camera parameter estimation; structure from motion; GPS;**

I. Introduction

Extrinsic camera parameter estimation for a moving video camera has been widely investigated and used for computer vision and virtual reality applications such as three-dimensional reconstruction [1], novel view generation [2] and augmented reality [3]. For these applications, Structure from Motion (SfM) technique has often been used. In SfM, most of the recent works employ the bundle adjustment that non-linearly minimizes the sum of reprojection errors [4], [5]. The SfM suffers from accumulative errors that cannot be resolved in principle if the method uses only images. This problem especially affects the applications that require the camera parameters in large-scale outdoor environments.

In order to reduce accumulative errors, some reference points or external sensors like GPS have been used in addition to video images [1], [6]–[10]. The former approach uses known 3D positions of reference points as prior knowledge about a target environment [6]. Although this method can estimate absolute camera positions without any other sensors, 3D measurement for the target environment is necessary and it requires much manual intervention. On the other hand, GPS and vision hybrid methods [1], [7]–[10] can also estimate absolute camera positions without pre-measurement. Thus, a GPS and vision hybrid is one of the promising solutions for extrinsic camera parameter estimation in large-scale outdoor environments. However, existing methods still have a problem with accuracy.

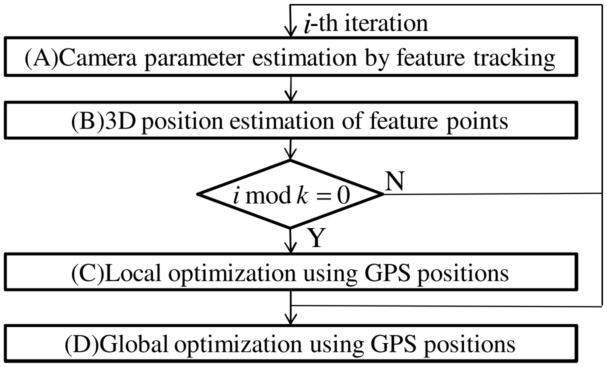
Conventional vision and GPS hybrid methods can be classified into Kalman filter based methods [1], [7], [8] and bundle adjustment based methods [9], [10]. Kalman filter based methods [1], [7], [8] tend to be employed for real-time applications because GPS and vision data can instantly be fused from the previous state and the current measurement. The problem of the Kalman filter is concerned with the difficulty of global optimization due to the sequential updating strategy of the filter design.

Bundle adjustment based methods [9], [10] use the energy function that is defined as the sum of reprojection errors and a penalty term of GPS. These methods can globally optimize the camera parameters by updating parameters so as to minimize the energy function. However, the accuracy of the estimated camera position largely depends on the actual accuracy of GPS because these methods assume that the GPS position error is very small [9] or normally distributed [10].

In this study, in order to obtain accurate extrinsic camera parameters even when the GPS positioning accuracy drops lower, we have employed a more simple and flexible penalty term for the bundle adjustment. This penalty term is designed by assuming that the true GPS position exists within a certain range from the observed GPS position and the size of the range depends on the GPS positioning accuracy.

II. Extrinsic camera parameter estimation considering GPS positioning accuracy

The proposed method basically follows the framework of vision and GPS hybrid methods [9], [10]. As shown in Figure 1, first, camera parameter estimation (A) and 3D position estimation of feature points (B) are repeated sequentially for each frame from the first frame to the last frame. In this repetition, at a constant frame interval k, the local optimization process (C) is applied to reduce accumulative errors. After estimating initial camera parameters by the processes (A) to (C), estimated parameters are globally refined (D). For all the processes, a common energy function is minimized. In the following, first, the energy function is



i-th iteration
(A)Camera parameter estimation by feature tracking
(B)3D position estimation of feature points
$i \bmod k = 0$
N
Y
(C)Local optimization using GPS positions
(D)Global optimization using GPS positions

Figure 1. Flow diagram of the proposed method.

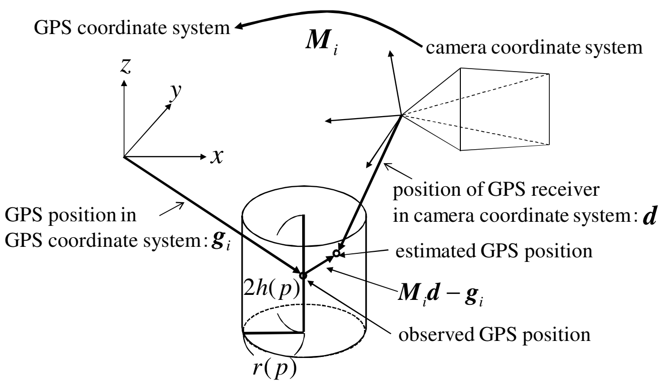


Figure 2. Coordinate system of GPS and camera.

defined by using the reprojection error and the penalty term for GPS positioning. Optimization process is then detailed.

A. *Energy function considering GPS positioning accuracy*

The energy function E is defined by using the reprojection error Φ_i and the penalty for GPS positioning Ψ_i for the i-th frame as follows:

$$E = \sum_{i \in \boldsymbol{F}} \Phi_i + \omega \sum_{i \in \boldsymbol{F}_g} \Psi_i, \tag{1}$$

where ω is a weight for GPS, $\boldsymbol{F}$ denotes a set of input frames and $\boldsymbol{F_g}$ denotes a set of frames in which GPS positioning data is obtained. It should be noted that GPS positioning data are acquired at 1Hz and that of video images are 15Hz or more. In the following, the energy associated with reprojection error Φ_i and the penalty energy for GPS positioning Ψ_i are detailed.

Reprojection error

Reprojection error is a squared distance between the detected 2D position of the feature point and the projected position of the corresponding 3D feature point in space. The reprojection error has often been used in SfM. In this study, the energy term associated with the reprojection error Φ_i is defined as follows:

$$\Phi_i = \frac{1}{|\boldsymbol{S}_i|} \sum_{j \in \boldsymbol{S}_i} \kappa_j (\boldsymbol{q}_{ij} - \hat{\boldsymbol{q}}_{ij})^2, \tag{2}$$

where $\boldsymbol{S}_i$ denotes a set of feature points detected in the i-th frame. κ_j represents the confidence of feature point j, which is computed as an inverse variance of the reprojection error [6]. $\boldsymbol{q}_{ij}$ and $\hat{\boldsymbol{q}}_{ij}$ represent the detected position of the feature j in the image i and the 2D projected position of the 3D position for feature j, respectively.

Penalty term for GPS positioning

In this study, the penalty term for GPS positioning is designed by assuming that the true GPS position exists within a certain range from the observed position of the GPS. Here, as shown in Figure 2, this range is set as a cylinder by assuming that GPS positioning error is independent of horizontal and altitude directions. Based on this assumption, the penalty term for GPS positioning Ψ_i is defined as follows:

$$\Psi_i = \left(\frac{1}{r(p)} \sqrt{x_i^2 + y_i^2} \right)^{2n} + \left(\frac{1}{h(p)} z_i \right)^{2n}, \tag{3}$$

$$\begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix} = \boldsymbol{M}_i \boldsymbol{d} - \boldsymbol{g}_i, \tag{4}$$

where $\boldsymbol{M}_i$ represents the transformation matrix from the camera coordinate system for the i-th frame to the GPS coordinate system, $\boldsymbol{d}$ represents the position of the GPS receiver in the camera coordinate system, $\boldsymbol{g}_i$ is the observed GPS position for the i-th frame in the GPS coordinate system. r and h are radius and half of the height of the cylinder, respectively, and they are functions of GPS positioning accuracy p. n is a given large number. This penalty term Ψ_i becomes very small if the estimated GPS position exists inside the cylinder. Otherwise, Ψ_i becomes very big.

B. *Optimization by minimizing the energy function*

In processes (C) and (D), in order to optimize the camera parameters and 3D positions of feature points, the energy function E is non-linearly minimized by the gradient method. The difference in processes (C) and (D) is the range of optimized frames. In process (C), in order to reduce accumulative errors in the sequential process, the parameters from the $(i - l)$-th frame to the current frame (i-th frame) are refined. In process (D), in order to globally optimize the camera parameters, the parameters for all the frames are refined.

III. Experiments

In order to validate the effectiveness of the proposed method, the result by the proposed method is compared with that of conventional methods [9], [10] quantitatively by using a real video sequence. This video sequence is captured by a video camera held by a walking person. Before comparing these methods, the cylinder size for the proposed method is determined by observing GPS positioning data at a fixed point for a long time.

A. Determination of error range

GPS positioning data are observed at a fixed point using RTK-GPS (TOPCON GR-3) for 5 hours. In this experiment, we use solution type of the RTK-GPS (RTK-fix, RTK-float) that is indicated from the GPS receiver as positioning accuracy p, because the size of the position error is generally depending on the solution type. Table I(a) shows maximum errors for each solution type. In this experiment, true position of the GPS is calculated as an average of all the RTK-fix solution data, which are generally more accurate than RTK-float solutions. Table I(b) shows maximum errors after removing the largest 5% of errors for each direction. From this table, it is obvious that there are outliers in the observed positions. In the following experiment, by assuming that outliers can be removed by checking the consistency between vision and GPS information, parameters of cylinder $r(p), h(p)$ are determined by using maximum errors shown in Table I(b).

It should be noted that, in the following experiment, the camera and GPS are moving between shutter timings of successive frames despite the synchronization of GPS and images is not done in sub-frame accuracy. In order to consider the camera motion between shutter timings, the size of the cylinder is actually set a little larger than values shown in Table I(b): 37mm and 5mm are added to the cylinder size of horizontal and altitude directions, respectively, by considering the person's walking distance in successive frames.

B. Quantitative evaluation

In this experiment, camera parameters are estimated for video images (720x480 pixel, 15 fps, progressive scan, 1110 frames) captured by a hand-held moving video camera (Sony DSR-PD-150) with a wide conversion lens (Sony VCL-HG0758). A GPS receiver (TOPCON GR-3) is attached to the camera and positioning data are acquired at 1Hz during video capture. In this experiment, all the GPS positioning data are acquired as RTK-fix solution data and they are used as the ground truth. From this data, lower accuracy GPS solutions (RTK-float) are generated in simulation by masking the GPS satellite data using the post process software (TOPCON Tools) and generated data are used as the input in this experiment. It should be noted that, in this experiment, we assume that the input data do not include the outliers.

The other conditions are as follows. Intrinsic camera parameters and the position of the GPS receiver in the camera coordinate system are calibrated in advance, and these parameters are fixed during video capture. The video frames and GPS input are manually synchronized. The weight ω in the error function E defined by Eq. (1) is set as 10^{-8} and n in the penalty term Ψ_i defined by Eq. (3) is set as 70. In the local optimization process (C), the frame interval k and the number of optimized frames l are set as 15 and 500, respectively.

The accuracy of estimated camera positions is compared with the following methods.

- Method A: The vision based method that does not use any GPS data.
- Method B: The conventional method [9]. That assumes GPS error follows a normal distribution and error is very small. The solution type of GPS data is not considered.
- Method C: The conventional method [10]. That treats error of GPS data as a normal distribution whose standard deviation is changed according to the solution type of GPS data.

In order to evaluate the camera positions by method A in the GPS coordinate system, some reference points of known 3D positions are given manually from the 1st to the 30th frame of the input video for method A. These reference points are also used for initializing the tracking process for methods B, C, and the proposed method.

Figures 3 and 4 show estimated GPS positions for horizontal and altitude directions, respectively. Table II shows position errors for each method. From these results, it is confirmed that method A is affected by accumulative errors. Method B is obviously affected by RTK-float solution data. Although the estimated error for method C is smaller than the error of method B that considers the accuracy of GPS, estimated positions are still biased to the RTK-float solutions. In the proposed method, camera positions are not clearly biased and it obtains the most accurate camera positions compared with other methods as shown in Table II.

Table I
MAXIMUM ERRORS FOR EACH SOLUTION TYPE [mm].

solution	(a) all data		(b) without outliers	
type	horizontal	altitude	horizontal	altitude
RTK-fix	793	677	29	41
RTK-float	12792	20424	3778	9504

Table II
COMPARISON OF POSITION ERRORS [mm].

Method	Average	Std.dev.	Max
A	1985	792	2929
B	1787	1497	4742
C	425	278	779
Proposed	241	145	495

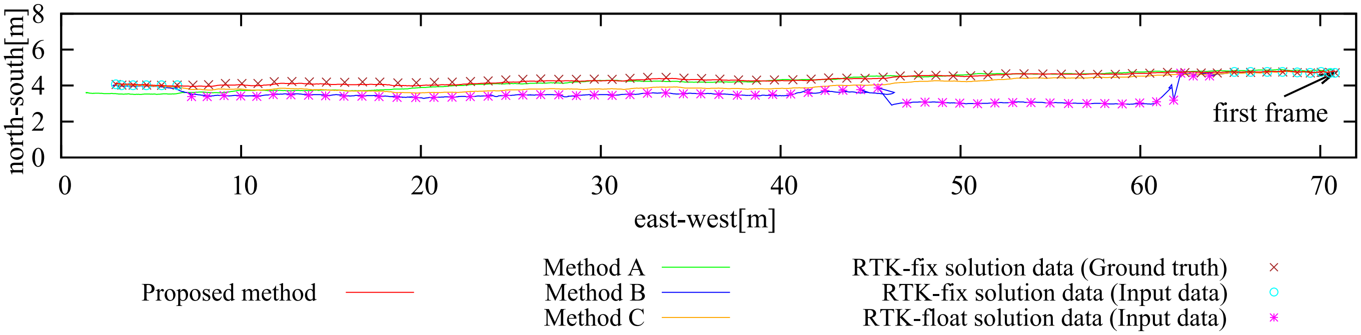


Figure 3. Estimated GPS positions (horizontal).

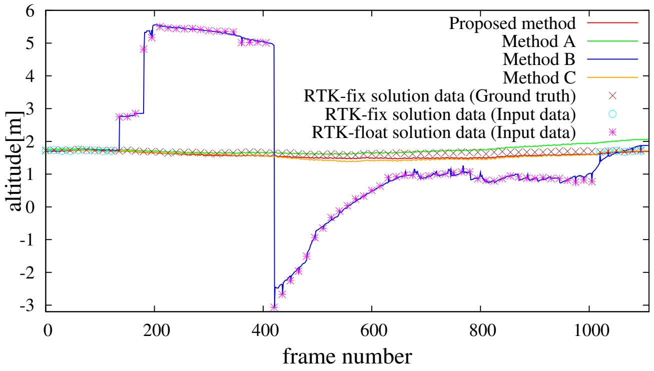


Figure 4. Estimated GPS positions (altitude).

IV. Conclusion

In this paper, we have proposed a vision and GPS hybrid method for accurately estimating extrinsic camera parameters even when the GPS positioning accuracy drops to a lower level by employing a simple and flexible penalty term for the bundle adjustment. In experiments, we have confirmed that the proposed method can obtain the most accurate camera positions compared with conventional methods. In future work, automatic outlier removal for GPS positioning will be incorporated with the proposed method.

Acknowledgment

This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology, Grant-in-Aid for Scientific Research (A), 19200016, and Strategic Information and Communications R&D Promotion Programme (SCOPE) of Ministry of Internal Affairs and Communications Japan.

References

[1] M. Pollefeys, D. Nistér, J. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. Kim, P. Merrell, et al.: “Detailed real-time urban 3D reconstruction from video,” Int. J. of Computer Vision, Vol. 78, No. 2-3, pp. 143–167, 2008.

[2] S. Knorr, M. Kunter and T. Sikora: “Super-resolution stereo- and multi-view synthesis from monocular video sequences,” Proc. Int. Conf. on 3-D Digital Imaging and Modeling, pp. 55–64, 2007.

[3] A. Stafford, W. Piekarski and B. H. Thomas: “Implementation of god-like interaction techniques for supporting collaboration between outdoor AR and indoor tabletop users,” Proc. Int. Symp. on Mixed and Augmented Reality, pp. 165–172, 2006.

[4] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops and R. Koch: “Visual modeling with a hand-held camera,” Int. J. of Computer Vision, Vol. 59, No. 3, pp. 207–232, 2004.

[5] 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.

[6] 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. J. of Computer Vision, Vol. 47, No. 1-3, pp. 119–129, 2002.

[7] M. Agrawal and K. Konolige: “Real-time localization in outdoor environments using stereo vision and inexpensive GPS,” Proc. Int. Conf. on Pattern Recognition, Vol. 3, pp. 1063–1068, 2006.

[8] D. Schleicher, L. M. Bergasa, M. Ocana, R. Barea and E. Lopez: “Real-time hierarchical GPS aided visual SLAM on urban environments,” Proc. Int. Conf. on Robotics and Automation, pp. 4381–4386, 2009.

[9] Y. Yokochi, S. Ikeda, T. Sato and N. Yokoya: “Extrinsic camera parameter estimation based-on feature tracking and GPS data,” Proc. Asian Conf. on Computer Vision, Vol. I, pp. 369–378, 2006.

[10] T. Anai, N. Fukaya, T. Sato, N. Yokoya and N. Kochi: “Exterior orientation method for video image sequences with considering RTK-GPS accuracy,” Proc. Int. Conf. on Optical 3-D Measurement Techniques, Vol. I, pp. 231–240, 2009.