THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

TECHNICAL REPORT OF IEICE.

Camera Recovery of an Omnidirectional Multi-camera System Using GPS Positions

Sei IKEDA†, Tomokazu SATO†, and Naokazu YOKOYA†

† Graduate School of Information Science, Nara Institute of Science and Technology
Takayama 8916–5, Ikoma, Nara, 630–0192, Japan
E-mail: †{sei-i,tomoka-s,yokoya}@is.naist.jp

Abstract This paper describes a novel method for estimating positions and postures of an omnidirectional multi–camera system from its multiple image sequences and sparse position data acquired by GPS. The proposed method is based on a structure-from-motion technique which is enhanced by using multiple image sequences as well as GPS position data. Moreover, the position data are also used to remove mis-tracked features. The proposed method allows us to estimate position and posture without accumulative errors and annoying effects due to moving objects in outdoor environments. The validity of the method is demonstrated through experiments using both synthetic and real outdoor scenes.

Key words Camera Recovery, Omnidirectional Multi-camera System, Global Positioning System,

1. Introduction

This paper addresses a problem of estimating position and posture of an omnidirectional multi-camera system (OMS) from its multiple image sequences and GPS position data. This problem is especially important in acquisition of images of outdoor environments and its geometric information. A typical instance is construction of a virtual reality environment using real images covering user's field of view [1]∼[3]. In this application, precise estimation of position and posture of the camera is required to show a user virtual objects superimposed seamlessly on presented real images in the environment. To use real images acquired outdoors for such kind of applications, camera recovery algorithm should possess two characteristics : (i) errors in estimating positions and postures are not accumulated; (ii) the algorithm is strong for non-rigid motion of video such as moving objects and incidence of light of the sun. Camera recovery problem can be theoretically solved only from image information by using feature tracking based methods [4], [5]. However, algorithms using only images acquired with a general single camera do not satisfy the characteristics above. Some kinds of sensors or some kinds of prior knowledge about surroundings should be used in order to satisfy them.

The simplest method is to measure the required information directly by using any appropriate sensors. Guven [6] integrated an RTK-GPS (Real Time Kinematic GPS), a magnetometer and a gyro sensor to obtain position and posture data without accumulative errors. Highly accurate calibration and synchronization among sensors is needed but this problem is hardly treated. Giving a kind of prior knowledge also resolves the accumulative error problem. As prior knowledge about surroundings, manually selected known 3D-positions [7], [8] (called feature landmarks) and manually created 3D models [9], [10] are sometimes used in addition to feature tracking. However, these are hardly applicable to large scale environments because much cost is required to obtain this information.

The most hopeful solution for the accumulative error problem is a combination of images and GPS positions [11]∼[13]. In this paper, we propose a camera recovery method for an omnidirectional multi-camera system (OMS) using both video sequences and GPS position data. The proposed method extends our previous works [11] for OMS whose field of view is wide and is hardly filled with moving objects. Our method is based on structure-from-motion with feature tracking and parameter optimization using GPS positions and video frames. In feature tracking, tentative parameters are estimated from GPS position data. They are used to avoid mismatching and to obtain correspondences of feature points among different cameras. In the optimization, a new error function defined by using GPS position data and re-projection error is minimized to determine position and posture parameters of the OMS. In our method, the following conditions are assumed: (i) OMS and GPS are correctly synchronized; (ii) the geometrical relation between all the cameras and the GPS receiver is always fixed; (iii) the distance between the GPS receiver and the representative camera of the OMS is known, and the direction of GPS receiver in camera coordinate system is unknown. In this paper, it is also assumed that OMS has been calibrated in advance and the intrinsic camera parameters (including lens distortion, focal length and aspect ratio) of each element camera are known.

In the remainder of this paper, we first formulate the camera recovery problem of an OMS using GPS positions in Section 2. The implementation of the proposed method is then described in Section 3. In Section 4, the validity of the method is demonstrated through experiments for both synthetic and real outdoor scenes. Finally, we give conclusion and future work in Section 5.

2. Formulation of Camera Recovery of Omnidirectional Multi-camera System

The goal of this study is to obtain position and posture parameters of an OMS and a direction of GPS receiver from camera when multiple video frames and GPS positions are given. The main topic described in this section is how to integrate GPS position data to the structure-from-motion problem. In the proposed method, the general structure-from-motion algorithm is enhanced to treat GPS position information.

In the general structure-from-motion algorithm, re-projection error that is observation error is minimized to obtain the parameters. First, we make it clear what the parameters are. Second, as one of observation errors in our problem, the re-projection error is briefly explained. The error concerning GPS, which is another observation error, is then modeled by using geometric relation between camera and GPS. Finally, we describe a new error function combining re-projection error and the error function concerning GPS.

2.1 Position and Posture Parameters of Omnidirectional Multi-camera System

Omnidirectional multi-camera system is constructed of a set of cameras such as Ladybug (Point Grey Research) which can obtain omnidirectional videos as shown in Figure 1. As mentioned in the previous section, we assume that position and posture relations among element cameras are known and fixed in this paper. The positions and postures of all the cameras can be expressed as a pair of position and posture of a representative camera. In the i-th frame, the transformation from the world coordinate system to the camera coordinate system of each element camera c can be expressed by the following matrix N_{ic} by using the transformation M_c from the world coordinate system of a calibration process to the camera coordinate system of the camera c $(= 0, 1, 2, 3...n)$.

Figure 1 A sampled frame of an acquired omnidirectional video. Right bottom is an image of vertical element camera. Others are horizontal ones.

$$\mathsf{N}_{ic} = \mathsf{M}_c(\mathsf{M}_0)^{-1}\mathsf{N}_{i0} = \begin{bmatrix} \mathsf{R}_{ic} & \mathbf{t}_{ic} \\ 0 & 1 \end{bmatrix}, \quad (1)$$

where $\mathbf{t}_{ic}$ and R_{ic} represent the translation and the rotation from the world coordinate system of the i-th frame to the camera coordinate system of the camera c. This problem is treated as estimation of position ($\mathsf{R}_i = \mathsf{R}_{i0}$) and posture ($\mathbf{t}_i = \mathbf{t}_{i0}$) of the representative camera (c=0).

2.2 Error Function for Optimization Process

Re-projection Error

Re-projection error is generally used for camera recovery based on feature tracking. The method for minimizing the sum of squared re-projection error is usually referred to as bundle adjustment. The re-projection error Φ_{ij} is defined as $|\mathbf{q}_{ij} - \hat{\mathbf{q}}_{ij}|$ for the feature j in the i-th frame, where $\hat{\mathbf{q}}$ represents the 2D projected position of the feature's 3D position and $\mathbf{q}$ represents the detected position of the feature in the image.

Error of GPS positions

Generally, if GPS positions and estimated parameters do not contain any errors, the following equation is satisfied in the i-th frame among the parameters (position $\mathbf{t}_i$, posture R_i), GPS position $\mathbf{g}_i$ and the position of GPS receiver $\mathbf{d}$ in the camera coordinate system.

$$\mathsf{R}_i\mathbf{g}_i + \mathbf{t}_i = \mathbf{d} \ (i \in \mathcal{F}), \quad (2)$$

where $\mathcal{F}$ denotes a set of frames in which GPS positions are obtained. However, if GPS position $\mathbf{g}_i$ and the parameters $\mathbf{t}_i$ and R_i contain some errors, we must introduce an error vector $\mathbf{n}_i$.

$$\mathsf{R}_i\mathbf{g}_i + \mathbf{t}_i = \mathbf{d} + \mathbf{n}_i. \quad (3)$$

In this paper, we introduce an error function Ψ_i related to

GPS receiver using the length of the error vector $\mathbf{n}$: $\Psi_i = |\mathbf{n}_i|$. This function means the distance between the measured position of the GPS receiver and the predicted one. Next, we describe a new error function E which is a combination of the error function Ψ_{ij} related to GPS receiver and the re-projection error Φ.

Error Function Concerning Feature and GPS

The new error function E is defined as follows:

$$E = \frac{\omega}{|\mathcal{F}|} \sum_{i \in \mathcal{F}} \Psi_i^2 + \frac{1}{\sum_i \sum_c |\mathcal{S}_{ic}|} \sum_i \sum_c \mu_i \sum_{j \in \mathcal{S}_{ic}} w_j \Phi_{ijc}^2, \qquad (4)$$

where ω means a weight for Ψ_i, and $\mathcal{S}_i$ denotes a set of feature points detected in the i-th frame. The coefficients μ_i and w_j mean the confidences for frame and feature, respectively. w_j represents the confidence coefficient of feature point j, which is computed as an inverse variance of re-projection error Φ_{ij}. The coefficient μ_i denotes the confidence of the i-th frame. Two terms in the right-hand side in Eq. (4) is normalized by $|\mathcal{F}|$ and $\sum_i \sum_c |\mathcal{S}_{ic}|$ so as to set ω as a constant value independent of the number of features and GPS positioning points.

3. Implementation of Camera Recovery Method of Omnidirectional Multi-camera System

The proposed method basically consists of feature tracking and optimization of camera parameters as shown in Figure 2. First, two processes of (A) feature tracking and (B) initial parameter estimation are performed in order. At constant frame intervals, the narrow range optimization process (C) is then carried out to reduce accumulative errors. Finally, estimated parameters are refined using many tracked feature points in the wide range optimization process (D). In the processes (C) and (D), a common optimization is performed. The difference in both processes is the range of optimized frames. In the process (C), the range of optimization is a small part of the input frames because future data cannot be treated in sequential process. On the other hand, in the process (D), a large number of frames are optimized and updated.

(A) Feature tracking :

The purpose of this process is to determine corresponding points between the current frame i and the previous frame $(i-1)$. The main strategy to avoid mismatching in this process is that feature points are detected at corners of edges by Harris operator [14] and detected feature points are tracked robustly with a RANSAC [15] approach.

In the first process (a), natural feature points are automatically detected by using the Harris operator for limiting feature position candidates in the images. In the next process (b), every feature in the $(i-1)$-th frame is tentatively matched with a candidate feature point in the i-th frame by using a standard template matching. In the third process (c), tentative parameters are then estimated by selecting correct matches using a RANSAC approach [15]. In the final process (d), every feature is re-tracked within a limited searching area in image frames of all the element cameras, which can be computed by the tentative parameters and 3D positions of the features.

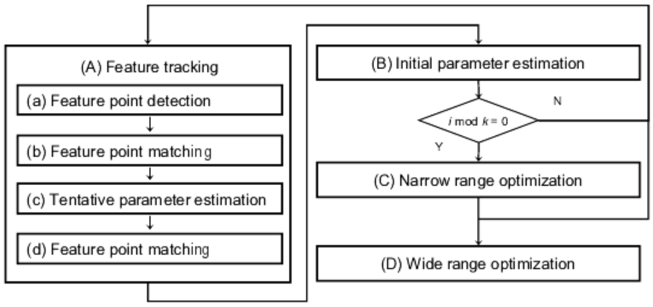


Figure 2 Overview of the proposed algorithm.

(B) Initial parameter estimation :

This processes computes 3D positions of feature points and position and posture parameters which minimize the sum of squared re-projection errors. In this process, the parameters of the current frame i are computed by using the tracked feature points. The error function E_{init} defined by Eq. (5) is minimized to optimize both the parameters $\mathbf{t}_i$ and R_i of all the frames and 3D positions of all the feature points.

$$E_{init} = \sum_{j \in \mathcal{S}_{ic}} w_j \Phi_{ijc}^2. \qquad (5)$$

(C) Narrow range optimization :

In this process, the frames from the $(i-(k+2l)+1)$-th to the current frame are used to refine the camera parameters from the $(i-(k+2l)+1)$ to the $(i-l)$-th frames, as illustrated in Figure 3. This process is designed to use feature points and GPS positions obtained in the frames around the updated frames. To reduce computational cost, this process is performed every k frames. Note that the estimation result is insensitive to the value of l if it is large enough. The constant l is set as tens of frames to use a sufficient number of feature points reconstructed in the process (B). The constant k is set as several frames, which is empirically given so as not to accumulate errors in the initial parameters estimated in the process (B).

It is difficult to obtain a global minimum solution because there are a large number of local minima in the error function E. In order to avoid this problem, we currently adopt a method to change the weight μ_i in the iteration of the op-

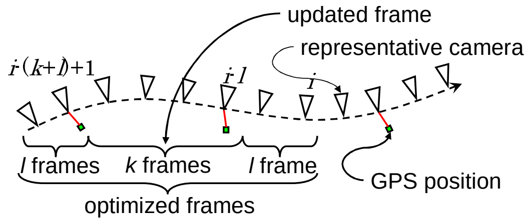


Figure 3 Optimization in the process (C).

timization, which is experimentally derived from computer simulations. In this method, the weight is changed whenever optimization process is converged. However, it should be note that there is a room for improvement because the present method is found just experimentally. This implementation is used in the next process (D)

(D) Wide range optimization :

The optimization in the process (C) dose not provide sufficient accuracy for a final output because it is performed for a part of frames and GPS positions for feedback to feature tracking process (A). The purpose of this process is to refine parameters by using tracked features and GPS positions in wide range frames. The algorithm of this process is the same as the narrow optimization process (C) when l and k are set as several hundred frames except that divided ranges are independent of each other.

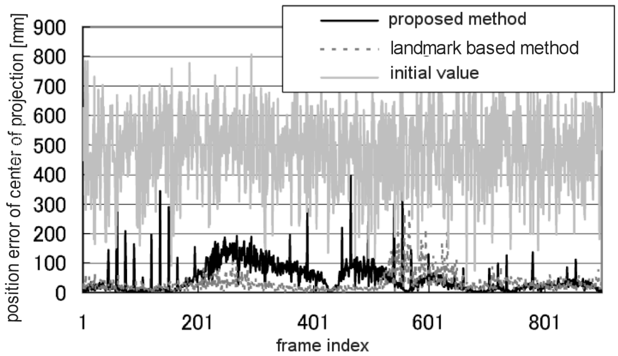
4. Experiment

This section describes experiments for both synthetic and real outdoor scenes. First, the experiment for synthetic data is carried out to evaluate the accuracy of the position and posture parameters of OMS estimated by the proposed method when the correspondences of feature points are given. The experiment for real data is then demonstrated to confirm the validity of the whole proposed method.

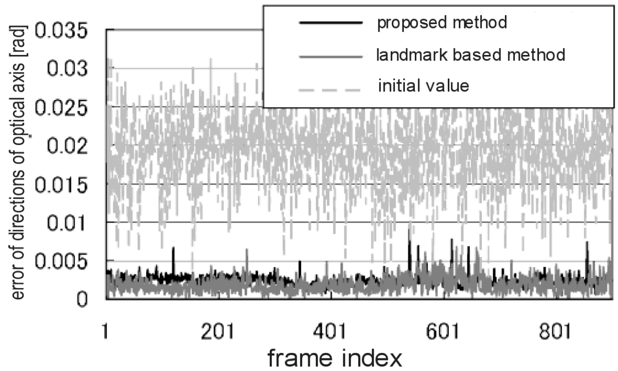
Note that some parameters used in the optimization process (C) and (D) were set as follows. The weight coefficient ω in the error function E defined by Eq. (4) was set as 10^{-9}. The weight μ_i of the corresponding frame was always set as 1.0, when a GPS position was obtained, When it was not obtained, 1.0 and 2.0 were alternately set as the weight μ_i whenever the optimization step was converged. In the narrow optimization process (C), we set the number of updated frames $k = 5$ and the number of optimized frames 49 ($l = 22$).

4.1 Synthetic Data

The purpose of this simulation is to evaluate the parameters $\mathbf{t}_i$ and R_i estimated in the wide optimization process (D). In addition, the validity of the proposed method is confirm by comparison with the conventional method [8]. We gave a point set as a virtual environment that was used to generate 2D feature positions in synthetic input images. The virtual camera takes 900 frames by moving in the virtual environment. The intrinsic parameters of the element camera in the virtual OMS are set the same as the real camera described in the next section. The position of GPS receiver in the representative camera coordinate system is set as (60,-150,250)[mm]. We added errors to input data as follows. The GPS positions with Gaussian noise (σ =30 mm) are given every 15 frames. The feature points are projected to the virtual camera, and detected with Gaussian noise (σ =1.6 pixel) and quantization error. The initial parameters are generated by adding Gaussian noise (position: $\sigma = 500$ mm, posture: $\sigma = 0.020$ rad) to the ground truth. In the compared method [8], all the frames is set as key frames in which more than 15 feature landmarks appear. The landmarks are given as feature points whose confidence coefficient is set as large enough, and the 2D positions of the landmarks in each frame are given without any errors.



(a) Errors of the estimated center of projection.



(b) Errors of estimated optical axes.

Figure 4 Error of the estimated position and posture.

Position and posture errors in the simulation result for the synthetic data are shown in Figure 4. In the compared method, the position error is 47.5 mm, and the postures error is 0.0019 rad on average. In the proposed method, the position error is 30.7 mm, and the posture error is 0.0023 rad on average.

These results indicate that the proposed method enables us to obtain position and posture parameters of OMS in the

Figure 5 Omnidirectional multi-camera system and GPS receiver.

same order of precision as the conventional method without any manual acquisitions of surrounding information. The difference in the accuracy between the proposed method and the compared one can be caused by the difference of the effect of the given absolute position information such as GPS positions and landmarks. Concretely, the reason why the posture errors are smaller than the position ones is that landmark position information obtained from images is more sensitive to the estimation of postures than GPS position information.

4. 2 Real Scene

The purpose of this experiment with real data is to confirm the validity of the proposed method which includes the feature tracking and the error models of feature point detection. In this section, we first describe the condition of the experiment. After that, two kinds of experimental results are shown.

In this experiment, we used Ladybug and a GPS receiver (Nikon LogPakII, horizontal accuracy ±3.0 cm, vertical accuracy ±4.0 cm) fixed on an electric mortar vehicle (see Figure 5). We acquired 7200 frames and GPS positions while the vehicle was moving 1.0km distance at 7.6km/h. The acquired frames and GPS positions were manually synchronized. Ladybug was calibrated by using the method developed in our previous work [16]. The distance between OMS and GPS receiver is 300 mm which is manually measured.

Figure 6 shows the estimated positions of Ladybug after the wide range optimization process (D). In this figure, the camera path is smoothly recovered and lies around the road. The match move using the estimated parameters is also demonstrated in Figure 7. The virtual cube in Figure 7 seem to be located at the same position in the real environment in most part of the input sequence. We have confirmed that estimated parameters do not contain fatal errors from these results.

In the last experiment, we confirm the result of optimization of divided sequences of the wide range optimization process (D). Figure 8 indicates the differences in estimated positions of the OMS between divided sequences and single se-

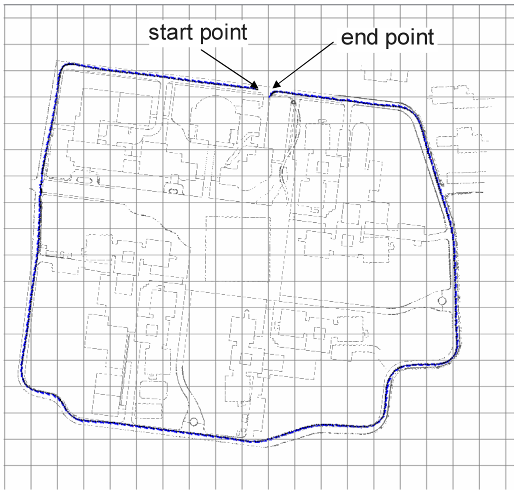


Figure 6 Horizontal trace of omnidirectional multi-camera system on the environment map.

quence. The ranges of divided sequences are shown in Figure 9. The difference of almost all of frames are smaller than the accuracy 30.7mm shown in the previous section 4. 1. This result indicates that we can deal with more long sequences by dividing the sequences if the range parameters k' and l' are set adequately.

5. Conclusion

In this paper, we have proposed a method to estimate positions and postures of an omnidirectional multi-camera system without accumulative errors from image data and coordinated GPS positions. In the proposed method, GPS position information is used for both feature tracking and optimization of position and posture parameters of the omnidirectional multi-camera system.

We have confirmed that the proposed method allows us to obtain extrinsic parameters in the same order precision as the conventional shape-from-motion method using a large number of landmarks if GPS positions are obtained well. We will investigate the case that the number of GPS positions including large errors is more than current experiments.

References

[1] S. Chen, “QuickTime VR: An image-based approach to virtual environment navigation,” Proc. SIGGRAPH ’95, pp.29–38, 1995.

[2] C. J. Taylor, “VideoPlus: A method for capturing the structure and appearance of immersive environment,” IEEE Trans. Visualization and Computer Graphics, vol.8, no.2, pp.171–182, 2002.

[3] M. Uyttendaele, A. Criminisi, S. B. Kang, S. Winder, R. Hartley, and R. Szeliski, “High-quality image-based interactive exploration of real-world environments,” IEEE Computer Graphics and Applications, 2004.

[4] A.W. Fitzgibbon and A. Zisserman, “Automatic camera recovery for closed or open image sequences,” Proc. 5th European Conf. on Computer Vision, pp.311 – 326, 1998.

[5] M. Pollefeys, R. Koch, M. Vergauwen, B. Deknuydt, and

Figure 7 Match move using estimated position and posture parameters of Ladybug.

L.V. Gool, "Three-dimentional scene reconstruction from images," Proc. SPIE, pp.215–226, 2000.

[6] S. Güven and S. Feiner, "Authoring 3D hypermedia for wearable augmented and virtual reality," Proc. 7th IEEE Int. Symp. on Wearable Computers, pp.118–126, 2003.

[7] A.J. Davison, "Real-time simultaneous localisation and mapping with a single camera," Proc. 9th IEEE Int. Conf. on Computer Vision, pp.1403–1410, 2003.

[8] 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 csamera," Int. Jour. of Computer Vision, vol.47, no.1-3, pp.119–129, 2002.

[9] A.I. Comport, É. Marchand, and F. Chaumette, "A real-time tracker for markerless augmented reality," Proc. 2nd ACM/IEEE Int. Symp. on Mixed and Augmented Reality, pp.36–45, 2003.

[10] L. Vacchetti, V. Lepetit, and P. Fua, "Combining edge and texture information for real-time accurate 3D camera tracking," Proc. 3rd IEEE and ACM Int. Symp. on Mixed and Augmented Reality, pp.48–57, 2004.

[11] 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, pp.369–378, 2006.

[12] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp.964–971, 2004.

[13] Z. Hu, U. Keiichi, H. LU, and F. Lamosa, "Fusion of vision, 3D gyro and GPS for camera dynamic registration," Proc. 17th Int. Conf. on Pattern Recognition, pp.351–354, 2004.

[14] C. Harris and M. Stephens, "A combined corner and edge detector," Proc. Alvey Vision Conf., pp.147–151, 1988.

[15] M.A. Fischler and R.C. Bolles, "Random sample consensus: A pradigm for model fitting with applications to image analysis and automated cartography," Communications of the ACM, vol.24, no.6, pp.381–395, 1981.

[16] S. Ikeda, T. Sato, and N. Yokoya, "High-resolution

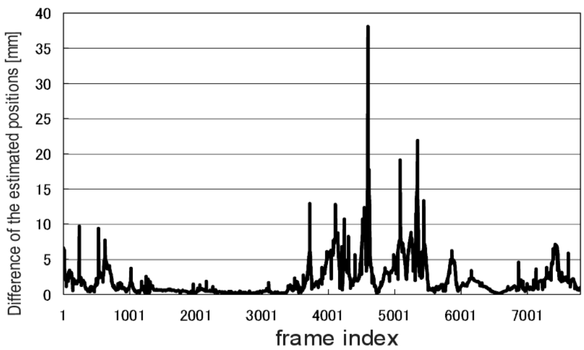


Figure 8 Difference of the estimaged positions from between divided sequences and single sequence.

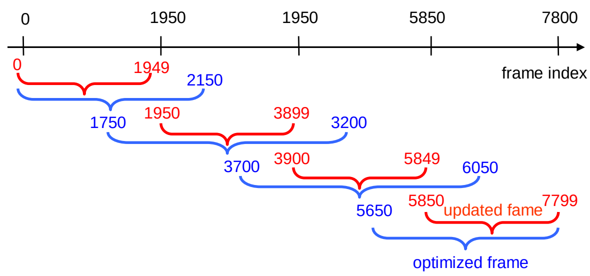


Figure 9 Range of optimization of divided sequence.

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.