Space-time Analysis of Spherical Projection Image

Shintaro Ono† Takeshi Mikami‡ Hiroshi Kawasaki‡ Katsushi Ikeuchi†

†The University of Tokyo
Ee–405, 4–6–1 Komaba, Meguro-ku,
Tokyo, JAPAN 153–8505
{onoshin,ki}@cvl.iis.u-tokyo.ac.jp

‡Saitama University
255 Shimo-ohkubo, Sakura-ku,
Saitama, JAPAN 338–8570
{mikami,kawasaki}@mm.ics.saitama-u.ac.jp

Abstract

In this paper, a novel analysis of space-time volume of spherical projection image is presented. So far, space-time analyses have been extensively conducted for various purposes, i.e. 3-D reconstruction, estimation of camera motion and novel view synthesis and most of them consider only a planer projection and a single camera. In contrast, we conducted analysis on spherical projection for multiple cameras. Since spherical projection does not change its appearance in relation to rotation around the origin of the sphere, extrinsic camera parameters and synchronous parameters of multiple video cameras can be simultaneously estimated by registering multiple space-time volumes of spherical projection, which can be easily achieved by block-matching technique. By using the parameters, multiple video images can be successfully integrated into single omni-directional images without distortions.

1 Introduction

In the past, space-time analyses have been extensively performed for the various purposes, such as the 3-D reconstruction of a scene, estimation of camera motion, novel view synthesis and so on. These research usually involved only a planer projection and a single camera at a time. In this paper, unlike the method, we presented a novel analysis of space-time volume, which involves spherical projection and the use of multiple cameras simultaneously.

A key notion proposed in this paper is representation of a spacetime image in a spherical solid and its analysis. In spherical space-time representation, the shape of a space-time volume acquired from a spherical projection image becomes invariant in relation to rotation around the origin of the sphere. Therefore, by registering multiple spacetime volumes in spherical space-time, extrinsic camera parameters and synchronous parameters of multiple video cameras are simultaneously acquired, and each camera image can be directly integrated with no distortion by using the parameters.

This paper is organized as follows. Section 2 describes related studies on space-time analysis, while section 3 explains various kinds of space-time representations including our proposed spherical projection. Following this discussion, we explain how to analyze multiple space-time volumes in section 4, and show the experimental results in section 5. Section 6 concludes our method.

2 Related works on space-time analysis

So far, extensive studies on space-time analysis of video camera images have been performed [2, 1, 10, 6, 7, 9]. However, most of the previous work involved only a planer projection and one camera at a time. Kawasaki *et al.*[4] proposed a space-time analysis for cylindrical projection and effectively retrieved 3D infromation. They also proposed the space-time analysis for multiple video cameras [5] to estimate synchronous parameters between them by analyzing multiple Epipolar Plane Images (EPI)[2, 1], where the cameras are mounted on the roof of a capturing vehicle.

However, their method presented several problems. First, position and orientation of multiple cameras must be known in advance and could not be refined. In the second place, they assumed a virtual projection plane common to every camera, and thus, non-uniformity of resolution in projected images arose, especially in the case of a camera direction is nearly parallel to the virtual plane.

In our proposed method, by using spherical projection for space-time analysis, extrinsic camera parameters and synchronous parameters of multiple video cameras can be comprehensively estimated, also providing a solution to the resolution problem.

3 Space-time volume of spherical projection

In this paper, we introduce a novel type of space-time representation, a spherical representation, which has unique

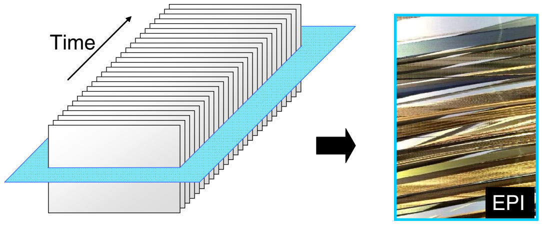

Figure 1. Planer projection of space-time and EPI.

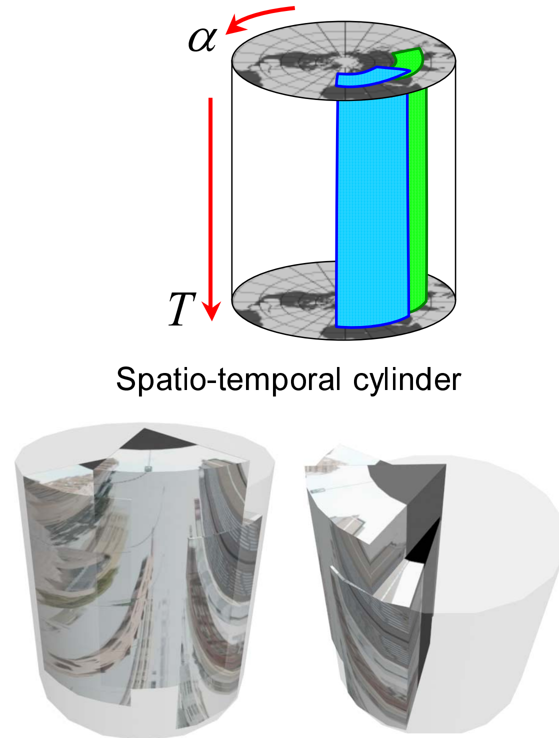

Figure 2. Cylindrical projection of spacetime.

features in regard to its invariancy. In advance of introducing it, we define simpler representations step-by-step and clarify features of our representation.

3.1 Planer projection

In this case, a space-time image is represented by a volume shaped as a rectangular solid as shown in Figure 1, which can be composed by simply accumulating sequential images along a temporal axis orthogonal to the images. It is known that an image that appears on the face of a horizontal cross-section of the volume becomes a so-called Epipolar Plane Image (EPI), if these sequential images are captured by a camera with horizontal movement. However, this space-time representation can not deal with more than two cameras at a time, except in the case of orthographic projection images. In order to consider associations between multiple video cameras, introduction of a common coordinate becomes necessary [5].

3.2 Cylindrical projection

Another considerable representation of space-time is to project a camera image to a virtual cylinder whose centroid is located at the optical center of the camera, and generate a volume as shown in Figure 2. Features of this cylindrical space-time volume are that an image similar to EPI appears on the face of cross-sections by a circle concentric with the cylinder, and a Panoramic View Image (PVI) appears on the face of cross-section by a radius of the cylinder [4].

In this representation, the shape of the volume is invariant in relation to rotation around the axis of the projection cylinder.

3.3 Spherical projection

Our novel representation of space-time proposed is a representation by a spherical projection. By defining the temporal axis proportional to radial direction of the sphere, we can represent space-time by a spherical solid. Figure 3-(a) shows the conceptual image of the representation. As in the case of the cylindrical projection, deformed patterns of EPI or PVI appear on the face of cross-section of the volume(Figure 3-(c) and (d)).

The most important feature of this representation is that shape of the spacetime volume is invariant in relation to any rotational shift around the origin of the sphere. Free parameters in this representation are any rotational elements around the origin (α, β and γ), in addition to time (T). In this paper, we call this representation of space-time volume a “spherical space-time representation.”

Taking advantage of these features, sequential images of multiple cameras captured under specific conditions can be spatially and temporally calibrated at once through 4-DOF (Degrees Of Freedom) volume-to-volume registration. Details are presented in the next section.

4 Analysis of multiple space-time volumes

In this section, we consider the interrelation of multiple spacetime volumes (STVs) . As described in the previous section, “spherical space-time representation” has four free parameters, and thus, STVs can be freely shifted along such parameters with its representation kept invariant. That is to say, if extrinsic parameters of multiple cameras, except for the above parameters, are known, and corresponding STVs

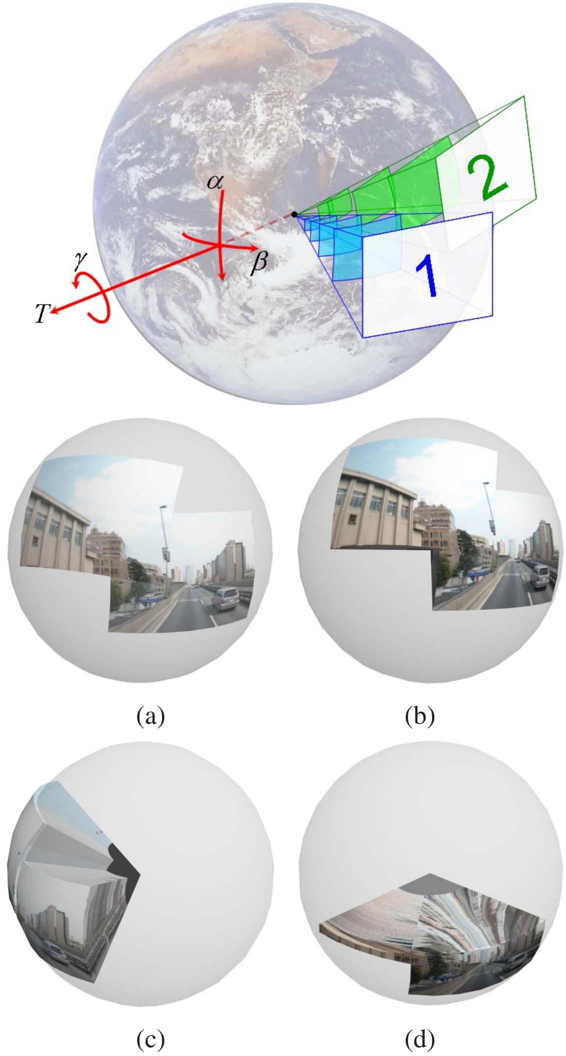

Figure 3. Concept(a) and examples(b-d) of spacetime sphere.

are converted to a united coordinate system, STVs can be registered by shifting them along free parameters.

4.1 Configuration of multiple cameras

First of all, we have to consider configuration of cameras where their extrinsic parameters, except free parameters, can become equal. Since a configuration with such a condition satisfied is proposed by Kawasaki *et al.*[5], we follow their configuration. As shown in Figure 4, multiple video cameras are arranged on the roof of a capturing vehicle in a line parallel to its moving direction. Although the optical centers of each camera are not spatially coordinated to the same point, they can be brought together at a different time in each camera. Therefore, by directing the optical axis of each camera radially, camera images looking in each direction can be captured per every sampling point on the running path of the vehicle. From here we can synthesize omni-directional images without distortions caused by disagreement of optical centers.

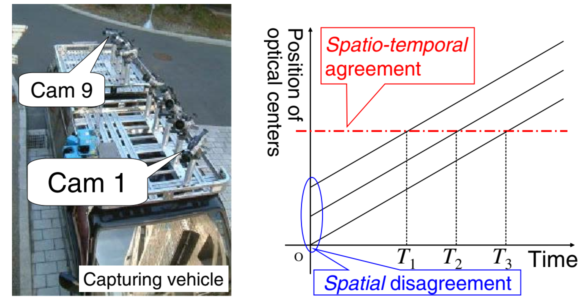

Figure 4. Configuration of multiple cameras.

For the sake of convenience, we assume the motion of the cameras is uniform-straight in this section. However, no limitation is imposed for camera motion in our proposed method itself.

4.2 spacetime parameters

Next, we clarify parameters to be estimated. When synthesizing multiple video streams into a single omni-directional video stream, camera parameters required for synthesis can be described as follows:

$$\begin{array}{llll} A_1, & R_1(T), & \boldsymbol{t}_1(T), & \delta_1(T) \equiv 0, \\ A_2, & R_2(T), & \boldsymbol{t}_2(T), & \delta_2(T), \\ \vdots & \vdots & \vdots & \vdots \\ A_N, & R_N(T), & \boldsymbol{t}_N(T), & \delta_N(T) \end{array} \tag{1}$$

where N is the number of cameras, T is time, A_n is the intrinsic matrix, R_n and $\boldsymbol{t}_n$ are the rotation matrix and transition vector, and δ_n is a synchronous parameter, i.e., the difference in time between cameras n and 1 when optical centers of each camera agree. In this paper, we call these parameters "spacetime parameters" of a video camera.

Meanwhile, in our camera configuration, some more constraints are added to these parameters. First, R_n and $\boldsymbol{t}_n$ are defined in reference to a coordinate system on the vehicle. Since we aim to synthesize omni-directional images seen from the vehicle coordinate system, they becomes invariant in relation to T. Second, regardless of the arrangement interval of the cameras shown in Figure 4, each camera passes through the point where camera No.1 existed at some previous time, and each camera image captured at that point can compose a distortion-free omni-directional image. This means that $\boldsymbol{t}_n$ is integrated inside synchronous parameter δ_n. δ_n also becomes invariant in relation to time, if movement of the vehicle can be regarded as uniform.

4.3 STV-STV registration

Next, we describe a concrete process of STV-STV registration. Hereafter, intrinsic parameter A_n is assumed to be estimated in advance by an existing method, and STV is successfully generated. Registration can be performed simply by shifting free parameters and searching an optimal point with the highest correlation between two STVs, F and G. The correlation can be evaluated by the following equation,

$$S(F,G) = \frac{\overline{F \cdot G} - \overline{F} \cdot \overline{G}}{\sqrt{\overline{F^2} - \overline{F}^2}\sqrt{\overline{G^2} - \overline{G}^2}} \qquad (2)$$

where $\overline{X}$ means to take an average of X inside intersected regions.

A positive feature of this method is that it does not require the searching of corresponding points, unlike conventional space-time analysis such as factorization. Registration can be robustly performed by a block-matching based process. Additionaly, little overlap for adjacent cameras is necessary in configuring cameras, since temporal information is fully utilized.

4.4 Image integration

Once spacetime parameters $A_n, R_n, \delta_n (n = 1, \ldots, N)$ are estimated, integrated video images can be acquired directly since the surface of a spacetime sphere or the cylinder itself at each time represents integrated images. See Figure 5(b) or 6 for an illustration of these images. Such image integration technique is an important research topic for both CV and CG[3, 8].

5 Experiments and results

To confirm the effectiveness of our proposed method, we carried out several experiments using real video images. Intrinsic camera parameters are assumed to be given as initial values. Video images are captured by driving along a downtown highway in our capturing vehicle, configured as shown in Figure 4.

As a result, matching three STVs in a spherical representation, is shown in Figure 5. In this case, STVs can be shifted to a latitudinal direction (β), in addition to a longitudinal and temporal direction (α, T). Three images each in Figure 5(a) are images on the surface of a spacetime sphere, estimeted to be correspond temporally as a result of 3-D block matching. As seen in Figure 5(b), camera images are successfully integrated into a panoramic image.

Figure 5(c) shows examples of a cross-section of matched STV, cut by a plane including the origin of the sphere. After the registration process, these cross-sections

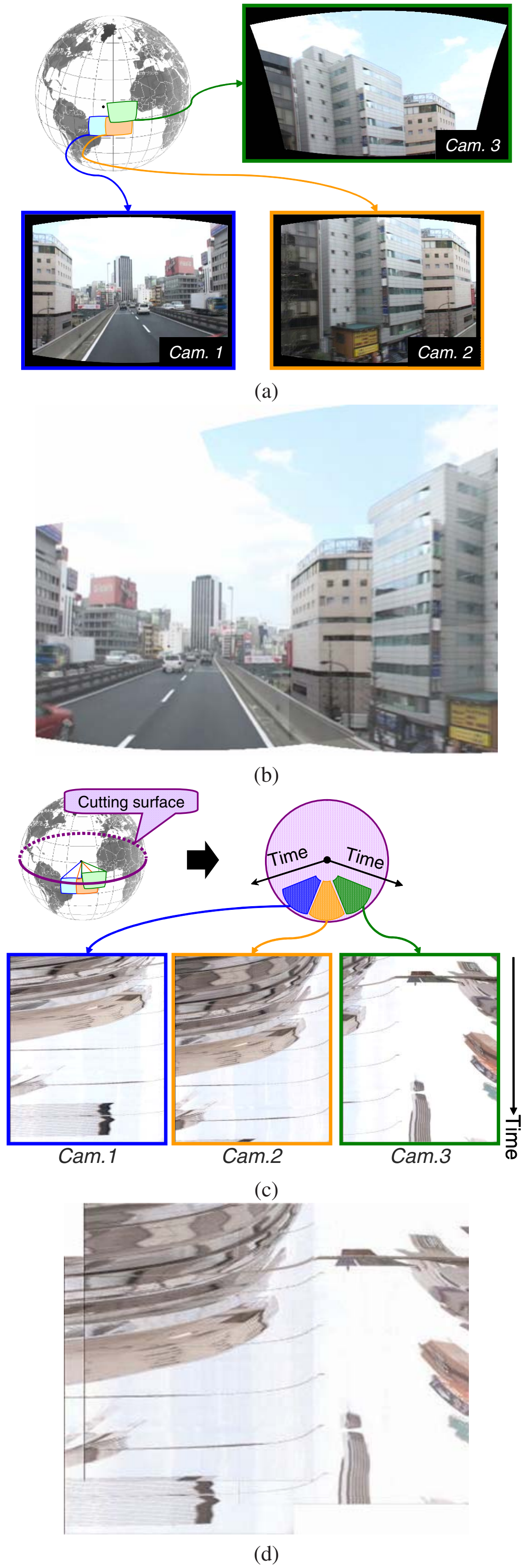

Figure 5. Result of STV registration.(a) input images, (b) stiched result and (c) cross section of STVs cut by a plane.

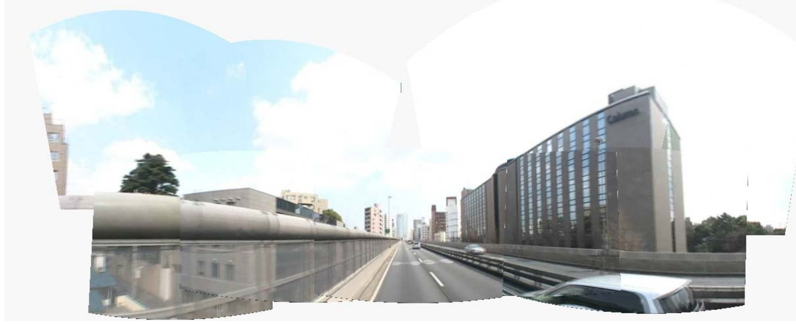

Figure 6. Registered result of STVs captured by nine video cameras.

were shifted to be integrated, which indicates that three videos are successfully integrated, achieving both spatial compliance and temporal compliance with each other.

Figure 6 shows another result of registering nine STVs.

6 Conclusion

In this paper, we have introduced a novel analysis for space-time volume of spherical projection to create distortion-free omni-directional images from multiple video cameras. Since spherical projection is consistent with any rotation around the optical center of a camera, extrinsic camera parameters and synchronous parameter of the muletiple video cameras can be estimated by correlation-based robust registration method.

By using these parameters, multiple camera images can be successfully integrated into a panoramic image without distortions, achieving both spatial and temporal compliance between all video streams. The effectiveness of our method can be confirmed through experiment using real videos.

References

[1] H. Baker and R. Bolles. Generalizing epipolar plane image analysis on the spatiotemporal surface. *Int.J.of Computer Vision*, 3:33–49, 1989. 1

[2] R. Bolles, H. Baker, and D. Marimont. Epipolar plane image analysis: an approach to determining structure from motion. In *Int. J. of Computer Vision*, volume 1, pages 7–55, 1987. 1

[3] M. Hirose and E. Takaaki. Building a virtual world from the real world. In *Proc. of International Symposium on Mixed Reality*, pages 183–197, 1999. 4

[4] H. Kawasaki, K. Ikeuchi, and M. Sakauchi. EPI analysis of omni-camera image. In *ICPR, IAPR 15th*, volume 1, pages 379–383, Barcelona, Spain, Sept. 2000. 1, 2

[5] H. Kawasaki, A. Miyamoto, Y. Ohsawa, S. Ono, and K. Ikeuchi. Multiple video camera calibration using EPI for city modeling. In *Asian Conf. Computer Vision*, volume 1, pages 569–574, Jan. 2004. 1, 2, 3

[6] A. Rav-Acha, Y. Pritch, D. Lischinski, and S. Peleg. Dynamosaics: Video mosaics with non-chronological time. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, June 2005. 1

[7] A. Rav-Acha, P. Yael, and P. Shmuel. Online registration of dynamic scenes using video extrapolation. In *10th International Conference on Computer Vision (ICCV 2005)*, Oct. 2005. 1

[8] T. Takahashi, H. Kawasaki, K. Ikeuchi, and M. Sakauchi. Arbitrary view position and direction rendering for large-scale scenes. In *CVPR*, volume 2, pages 296–303, 2000. 4

[9] Y. Wexler and D. Simakov. Space-time scene manifolds. In *International Coference on Computer Vision (ICCV 2005)*, Oct. 2005. 1

[10] Z. Zhu, G. Xu, and X. Lin. Panoramic EPI generation and analysis of video from a moving platform with vibration. In *CVPR99*, pages 531–537, 1999. 1