2005년도 대한전자공학회 추계종합학술대회 제28권 제2호

BAMA 알고리즘을 이용한 전방위 파노라믹 모자이크 생성

유호준*, 김성대**
한국과학기술원 전기전산학과 전기 및 전자전공

Generation of full panoramic mosaics using bidirectional alignment with multi-anchor algorithm

Ho June Leu*, Seong Dae Kim**
Division of EE, Department of EECS,
Korea Advanced Institute of Science and Technology
E-mail : *terius@sdvision.kaist.ac.kr, **sdkim@sdvision.kaist.ac.kr

Abstract

The acquisition of a panoramic image has been popular in recent years due to its large number of applications. In this paper, we deal with generation of a full panoramic mosaic. When it mosaicing consecutive frames, a simple alignment method may lead to local and global alignment errors. These errors can make a full-view panoramic mosaic to be cracked. In order to solve this problem, we propose the bidirectional alignment with multi-anchor (BAMA) algorithm with a basic local alignment formula. Experimental results show that we can create full panoramic mosaics stably through the proposed algorithm.

I. Introduction

A panoramic image provides a larger field of view for the taken scenes or environments than the field of view of the common single image. It also provides not only the augmented reality but immersion. Such panoramic images can be applied to the photo-realistic virtual reality, object segmentation, target tracking, and content based video codec, etc.

The acquisition of a panoramic image has been attracted tremendous attention from many researchers due to these various applications. The acquisition methods are mainly classified into two approaches. The first approach is to enhance the field of view of the lens. It has the advantage which can generate the panoramic image easily through the simple stitching algorithm using particular lenses (e.g. fish-eye lens [1], omni-directional lens [2], etc.). However, this method requires the expensive system and has the limitation of resolution. Second, there is the approach that it composes many frames with a narrow field of view from conventional lenses. This approach can construct a high resolution panoramic mosaic with cheap equipments. On the other hand, it needs complex alignment algorithms for preventing errors because one cannot take the full scene simultaneously. The sources of the trouble are the variation of the optical center due to handheld camera [3], the existence of moving object during taking scenes [4] and other variations (e.g. illumination change, auto exposure [5], and lens distortion [6], etc.).

These factors lead to local and global alignment errors. A local alignment error occurs during estimating the correspondences between neighboring frames, and a global error is revealed when the local error is accumulated. Many solutions for the local error were developed in recent years. However, many efforts for a global alignment error problem were not preceded. Several methods were presented for minimizing a global error [7-10]. In some cases like creating a full-view panoramic mosaic, previous approaches for minimizing a global error were not enough in order to remove any flaw (e.g. an end-to-end misalignment and so on).

Therefore, we propose the robust algorithm which can generate a flawless full panoramic mosaic. In section 2, we present the basic procedure to estimate the relationship between consecutive frames in a unidirectional way. In the next section, we propose BAMA algorithm to remove any flaw. We prove the efficiency of the proposed algorithm in section 4 and section 5 summarizes and concludes the paper.

II. Unidirectional Alignment

We estimate the projective transformation of each frame with cylindrical coordinates based on two assumptions in section 2.1.1. The simple concatenation of the transformations between neighboring frames may lead to a serious global alignment error. Therefore, it requires the strategy for estimating the transformation.

2.1 Local alignment

The relationship between each pair of consecutive frames having arbitrary camera motion is described as in eq. (1).

$$\mathbf{x}_{i+1} = \mathbf{K}_{i+1}\mathbf{R}_{i+1,i}\mathbf{K}_i^{-1}\mathbf{x}_i + d(\mathbf{x}_i)\mathbf{K}_{i+1}\mathbf{t}_{i+1,i} \qquad (1)$$

where $\mathbf{x}_{i+1}, \mathbf{x}_i$ are corresponding points in frame i+1 and i with Cartesian coordinates and $\mathbf{K}_{i+1}, \mathbf{K}_i$ are the intrinsic camera matrices of two frames. $\mathbf{R}_{i+1,i}, \mathbf{t}_{i+1,i}$ are the relative extrinsic camera matrices, rotation and translation, respectively and $d(\mathbf{x}_i)$ is the projective depth of the point $\mathbf{x}_i$.

2.1.1 Assumption

Our local alignment method needs two assumptions. First, we assume that there is scarcely any translation of the optical center between adjacent two frames.

$$\mathbf{t}_{i+1,i} \approx \mathbf{0}$$
$$\mathbf{x}_{i+1} \sim \mathbf{K}_{i+1}\mathbf{R}_{i+1,i}\mathbf{K}_i^{-1}\mathbf{x}_i = \mathbf{H}_{i+1,i}\mathbf{x}_i \qquad (2)$$

That is, the correspondences between neighboring frames approximate to the infinite homography, $\mathbf{H}_{i+1,i}$, as shown in eq. (2).

Next, if the narrow region of each frame from its center line is taken, the relationship between two frames in cylindrical coordinates can be approximated to a linear equation as in eq. (3). Eq. (4) shows that not only in Cartesian coordinates the relationship is satisfied but in cylindrical coordinates is. This is because there is the high similarity between neighboring frames.

$$\mathbf{x}_i = \begin{pmatrix} f_i \tan\theta_i \\ f_i v_i \sec\theta_i \\ 1 \end{pmatrix} \approx \begin{pmatrix} f_i\theta_i \\ f_i v_i \\ 1 \end{pmatrix} = \boldsymbol{\omega}_i \qquad (3)$$

where $\boldsymbol{\omega}_i$ is the cylindrical coordinate of ith frame and θ_i, v_i are panning angle and tilting respectively. f_i is the focal length.

$$\boldsymbol{\omega}_{i+1} \sim \mathbf{K}_{i+1}\mathbf{R}_{i+1,i}\mathbf{K}_i^{-1}\boldsymbol{\omega}_i = \mathbf{H}_{i+1,i}\boldsymbol{\omega}_i \qquad (4)$$

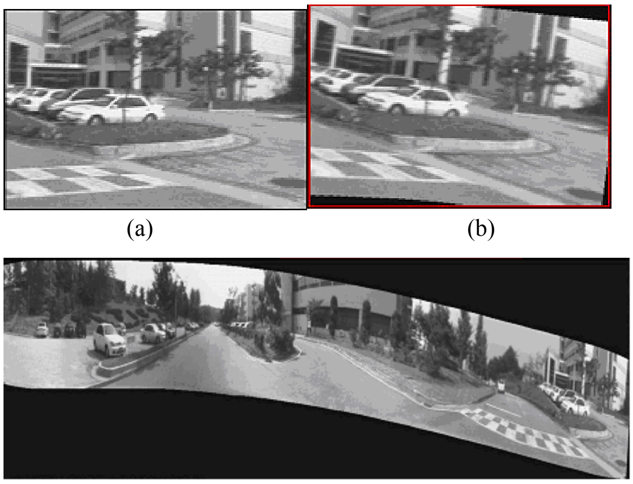
Therefore, if the narrow band of each input frame is taken, we can also estimate the homography with cylindrical coordinates as in eq. (4).

2.2. Estimation strategy

As mentioned in section 1, in general, although local homographies provide good correspondences between neighboring frames, the simple concatenation of these correspondences may still distort the panoramic mosaic. This is illustrated by the example in fig. 1.

In order to reduce or minimize the global alignment error, some algorithms were developed. We use the simple Smolic's method [10] among them as introduced in fig. 2. The cost function to minimize the error is described as eq. (5).

$$\varepsilon = \sum_{\boldsymbol{\omega}_i \in R_i} \left(I_{i+1}\left(\mathrm{W}\left(\boldsymbol{\omega}_i; \mathbf{H}_{i+1,i}\right)\right) - \tilde{I}_i\left(\boldsymbol{\omega}_i\right) \right)^2 \qquad (5)$$



(a) (b) (c)

Fig. 1. Error accumulation: (a) the original 211th frame, (b) the indexed 211th frame from the mosaic (c), (c) the generated cylindrical mosaic using frame-to-frame alignment

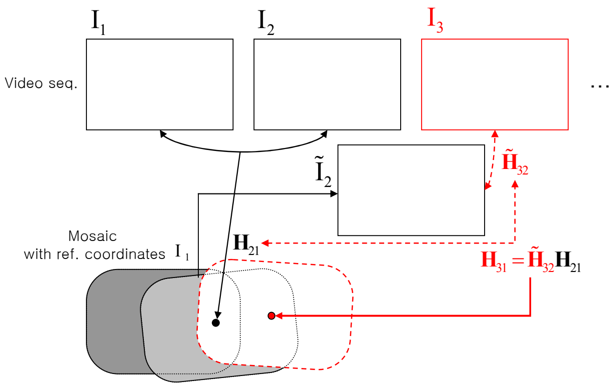


Fig. 2. Estimation strategy

where $\mathrm{W}(\boldsymbol{\omega}_i; \mathbf{H}_{i+1,i}) \equiv \boldsymbol{\omega}_{i+1}$ is the warping function in cylindrical coordinates and $\tilde{I}_i$ is the intensity value of ith frame to be indexed from reference coordinates. In this paper, Gauss-Newton method which is one of the nonlinear iterative optimization algorithms is used [11].

III. Bidirectional Alignment

The complete full panoramic mosaic cannot be obtained from a unidirectional alignment only. It needs an end-to-end alignment. In this chapter, we propose the BAMA algorithm for solving this problem.

3.1 Bidirectional Alignment with Multi-Anchor (BAMA) algorithm

BAMA algorithm estimates the homography of each frame spatio-temporally and creates full panoramic mosaics. Here, an anchor is defined as the pre-decided frame with the reliable homography. In 1-anchor case, the anchor is the frame with reference coordinates. It means the first frame in a sequence usually. If two anchors are used, the first anchor is the same as the above case and the second is defined as the frame which is the closest in the spatial domain and is the farthest in the time domain from the other anchor (e.g. the first and the last frames in a panning sequence). We can consider that the estimated homography for the second anchor is true because the second is the nearest to the first in the spatial domain. If 3 anchors are used, as illustrated in fig. 3, two anchors are the same as 2-anchor case. The last is defined as the farthest frame in the spatial domain from the others (e.g. a center frame in a panning sequence). The homography for the third anchor can be decided in various ways. Here, we assume that the average of the estimated homographies in bidirectional ways is the true value for the third.

As shown in fig. 3, we can estimate two homographies for all frames in forward and backward directions. The final homography for i^{th} frame is decided by the weighted average of two homographies as in eq. (6).

$$\mathbf{H}_{i,1}^{k}(\alpha_i)=\alpha_i\mathbf{H}_{i,1}^{F_k}+(1-\alpha_i)\mathbf{H}_{i,1}^{B_k}$$
$$0\le\alpha_i\le 1,\ k=0,1,\cdots,N-1 \qquad (6)$$

where $\mathbf{H}_{i,1}^{F_k}$ is the forward estimated homography between i^{th} frame and the reference frame and $\mathbf{H}_{i,1}^{B_k}$ is the backward estimated homography.

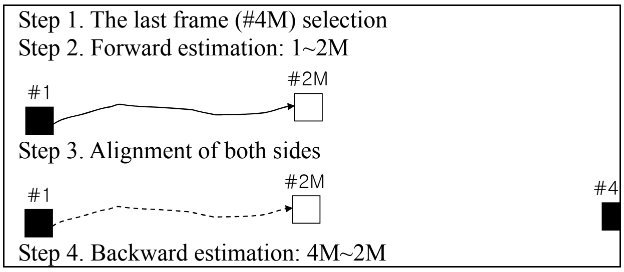
In general, the farther a frame is from the starting anchor, the larger α_i is. The initial value of α_i is as in eq. (7).

$$\alpha_i=\frac{2\mathrm{M}(\mathrm{k}+1)-\mathrm{i}-1}{2(\mathrm{M}-1)} \qquad (7)$$

The final α_i is decided by minimizing the cost function in eq. (8).

$$\varepsilon=\sum_{\omega_1\in R_1}\left(I_i\left(\mathrm{W}\left(\omega_1;\mathbf{H}_{i,1}^{k}(\alpha_i)\right)\right)-I_1(\omega_1)\right)^2 \qquad (8)$$

Total number of the computation for all frames is 4M if it is 1-anchor and is 8M if it has 2 or 3 anchors.



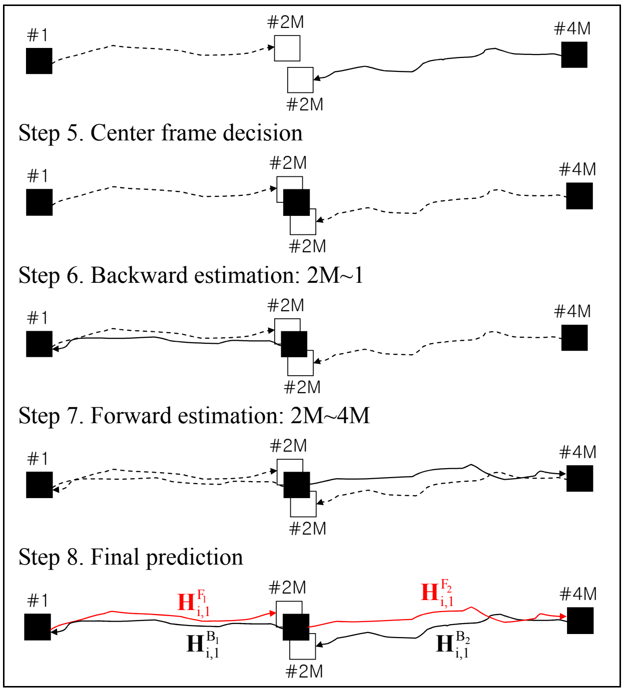


Fig. 3. BAMA algorithm (3 anchors)

IV. Experimental Results

All test sequences are taken from a digital camcorder. All images have 360x240 pixels. A quarter of its width from the center line of the image is used.

Fig. 4 is the result after unidirectional alignment. In all cases, an end-to-end was not aligned yet. The sequence in fig. 4 (a) was obtained from a camcorder on a tripod. The remained was get from a handheld camcorder. In fig. 4 (a), we can see that error accumulation is nearly occurred but we can see the large end-to-end misalignment within results in fig. 4 (b)-(d). Especially, error accumulation is serious in fig 4 (d).

The final results using BAMA algorithm (3 anchors) is shown in fig. 5 (a)-(d). We can see our algorithm obtains panoramic mosaics with high quality without any misalignment in all cases.

Fig. 4. (a/b/c/d) Results after unidirectional alignment: (a) the cylindrical mosaic using a tripod, (b)-(d) cylindrical mosaics with hand-held camcorder

Fig. 5. (a/b/c/d) Final results after BAMA algorithm: (a) the cylindrical mosaic using a tripod, (b)-(d) cylindrical mosaics with hand-held camcorder

V. Conclusion

We proposed the simple and robust BAMA algorithm with a basic local alignment formula which can generate full panoramic mosaics from images with handheld camera. In the presented algorithm, homographies between each frame and the reference frame were estimated from bi-directional alignment and then they were optimized by minimizing the cost function. Therefore, although it makes the error owing to various sources, we can create panoramic mosaics with good quality without distorting the geometric structure of them.

References

[1] Y. Xiong, and K. Turkowski, “Creating Image-based VR Using a Self-Calibration Fisheye Lens,” *IEEE Computer Society Conference on CVPR'97*, pp. 237-243, 1997.

[2] S. K. Nayar, “Catadioptric Omnidirectional Camera,” *IEEE Computer Society Conference on CVPR‘97*, pp. 482-488, 1997.

[3] S. Peleg and J. Herman, “Panoramic mosaics by manifold projection,” *IEEE Computer Society Conference on CVPR’97,* pp. 338-343, 1997.

[4] J. Davis, “Mosaics of Scenes with Moving Objects,” *IEEE Computer Society Conference on CVPR’98,* pp. 354-360, 1998.

[5] M. Frank, “Simultaneous Homographic and Comparametric Alignment of Multiple Exposure-Adjusted Pictures of the Same Scene,” *IEEE Transactions on Image Processing*, pp. 1485-1494, December 2003.

[6] H. S. Sawhney and R. Kumar, “True multi-image alignment and its application to mosaicing and lens distortion correction,” *IEEE Computer Society Conference on CVPR’97*, pp. 450-456, 1997.

[7] D. Capel and A. Zisserman, “Automated mosaicing with super-resolution zoom,” *IEEE Computer Society Conference on CVPR’98,* pp. 885-891, 1998.

[8] H. Y. Shum and R. Szeliski, “Construction of panoramic image mosaics with global and local alignment,” *International Journal of Computer Vision*, 36(2), pp101-130, 2000.

[9] R. Marzotto, A. Fusiello and V. Murino, “High Resolution Video Mosaicing with Global Alignment,” *IEEE Computer Society Conference on CVPR’04,* pp. I-692-I-698, 2004.

[10] A. Smolic and T. Sikora, “Long-Term Global Motion Estimation and Its Application for Sprite Coding, Content Description, and Segmentation,” *IEEE Transactions on Circuits and Systems for Video Technology,* Vol. 9, pp. 1227-1242, 1999.

[11] S. Baker and I. Matthews, “Lucas-Kanade 20 Years On: A Unifying Framework,” *International Journal of Computer Vision,* 56(3), pp. 221-255, 2004.