IEICE TRANS. INF. & SYST., VOL.E86–D, NO.7 JULY 2003

LETTER

Fast Algorithm for Aligning Images Having Large Displacements

JunWei HSIEH[†] *and* **Cheng-Chin CHIANG**[†], *Nonmembers*

SUMMARY This paper presents an edge alignment method for stitching images when they have large displacements and light changes. First, without building any correspondences, the proposed method predicts all possible translation solutions by examining the consistency between edge positions. Then, the best solution can be obtained from the set of possible translations by a verification process. The proposed method has better capabilities to stitch images when they have large light changes and displacements. Since the method doesn't require building any correspondences or involve any optimization process, it performs more efficiently than other correlation techniques like feature-matching or phase-correlation approaches. Due to its simplicity and efficiency, different images can be very quickly aligned (less than 0.1 seconds) with the proposed method. Experimental results are provided to verify the superiority of the proposed method.

key words: *image stitching, video surveillance, video indexing, image mosaics*

1. Introduction

Image alignment is a technique to estimate camera motion from images and then stitch them together. This technique has been successfully applied to many different applications like video compression [1], video indexing [2], [3], video object tracking, or mosaic construction [5]–[8], [10]. For example, Shum and Szeliski [7], [8] proposed methods to stitch a set of images together for constructing a panorama. In addition, Irani and Anandan [2] used this technique to construct different video mosaics for indexing video contents. Another important application of this technique is to extract interesting objects from images for video surveillance. For most methods in this field, an affine camera model is used to approximate possible motions between pairs of consecutive frames. Then, the wanted motion parameters can be obtained with a nonlinear optimization process by minimizing the discrepancy in intensities between images. However, this technique requires that the displacement between images be small; otherwise, the method will become trapped in a local minimum. For avoiding this problem, two common methods are adopted to estimate the desired displacement in advance, i.e., the frequency-based method and the feature-based one. For the first approach, Kuglin and Hines [4] presented a phase-correlation approach to estimate the displacement between images by transforming images into frequency domain. However, the technique requires the images having large overlapping and small intensity changes. As to the second approach, Sawhney and Ayer [1] proposed a feature matching approach to estimate dominant camera motions from images. In addition, Zoghlami et al. [9] proposed a corner matching approach to compute motion transformations from pairs of images. However, the success of this technique depends strongly on the establishment of correct correspondences and this work is still rigorously challenged in the field of computer vision [12].

In this paper, we propose an edge-based method for aligning images by examining the consistency between edge positions. The method first uses edge operators to extract various vertical and horizontal edges from images. Then, without building any correspondences, the paper tries to find all possible translations by examining the similarity between extracted edge positions. Then, with a verification technique, the best solution can be determined from the set of possible translations. Three advantages can be gained from this approach. First, the proposed method estimates possible translations based only on edge positions. Edge position is more robust to keep invariant than image intensity if light condition changes. Therefore, the proposed method can perform quite robustly when light changes. Second, since the method estimates image displacements without involving any optimization processes or building any correspondences, the proposed method performs more efficiently than other methods like feature matching [2] or constrained optimization [6]. Third, since the method is very simple, it is easily to be hardware-implemented for industrial inspection applications. Although the method deals with only the translation problem, the result can provide a good initialization for further optimization and image analysis in various applications like video retrieval, surveillance, and mosaic construction. Experimental results show the proposed method offers great improvements in terms of accuracy, robustness, and stability in image alignment.

The rest of the paper is organized as follows. In the next section, details of the proposed method including edge detection, edge alignment, and parameter estimation are described. Section 3 reports the experimental

Manuscript received September 17, 2002.
Manuscript revised December 20, 2002.

[†]The authors are with the Department of Electrical Engineering, Yuan Ze University, 135 Yuan-Tung Road, Chung-Li 320, Taiwan.

results. Then, conclusions will be presented in Sect. 4.

2. Fast Image Alignment Algorithm

In this paper, the camera is assumed to have only translational motions. If the rotation change is large, the rotation angle can be estimated and compensated by our previous work [11]. Based on this assumption, this paper presents a novel edge-based method to estimate desired translations between consecutive images by image alignment. First, some useful horizontal and vertical edges are extracted by using two specially designed edge operators. Then, all possible translation candidates are estimated with a novel edge alignment method. The overall flowchart is shown in Fig. 1. In what follows, details of the proposed algorithm are described.

Let $g_x(p)$ denote the gradient of a pixel p in the x direction of an image I, i.e.,

$$g_x(p(i,j)) = |I(p(i+1,j)) - I(p(i-1,j))|,$$

where $I(p)$ is the intensity of p. In addition, let $S_g(i)$ denote the average sum of $g_x(p)$ obtained by accumulating $g_x(p)$ along pixels in the ith column, i.e,

$$S_g(i) = \frac{1}{H}\sum_j |I(p(i+1,j)) - I(p(i-1,j))|,$$

where H is the height of image I. If $S_g(i)$ is larger than a threshold, i.e., 15, the ith column is considered to have a vertical edge. After checking all pixels in image I column by column, a set of vertical edges can be found.

Assume I_a and I_b are two images with the same dimension $W \times H$ and prepared to be stitched and shown in Fig. 2 (a) and (b), respectively. Through the above vertical edge detector, the positions of vertical edges in I_a and I_b can be obtained as $P_a^v = (100, 115, 180, 200, \ldots, 470)$ and $P_b^v = (20, 35, 100, 120, \ldots, 390)$, respectively. If I_a and I_b come from the same static scene, there should exist an offset d_x such that $P_a^v(i) = P_b^v(j) + d_x$ and the corresponding relation between i and j is one-to-one. Then, d_x is the desired solution of horizontal translation between I_a and I_b, i.e., $d_x = 80$. Based on this idea, we want to present a novel method to estimate the translation parameters without building any correspondences or involving any optimization processes.

Before describing the proposed method, we shall know in some cases due to noise, some edges will be lost or undetected. The lost or undetected edges will lead to that the relations between P_a^v and P_b^v are no longer one-to-one. In order to deal with this problem, we define a function $d_v(i,k)$ to measure the distance of a position $P_a^v(i)$ to the translation solution k as:

$$d_v(i,k) = \min_{1 \le j \le N_b^v} |P_a^v(i) - k - P_b^v(j)|, \quad (1)$$

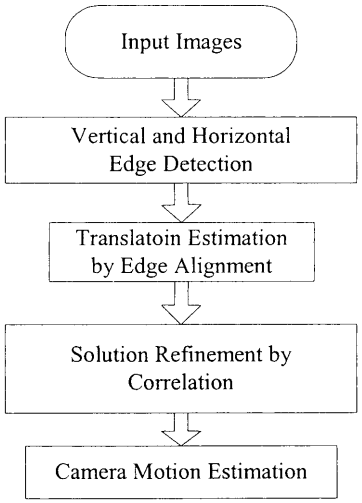
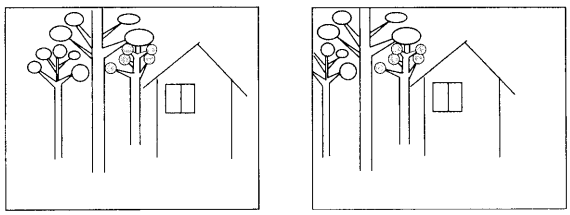


Fig. 1 Flowchart of the proposed method.



(a) (b)

Fig. 2 Edge results of two images.

where N_b^v is the number of elements in P_b^v. Let T_d denote a threshold and set to 4. Given a number k, we want to determine the number N_p^v of elements in P_a^v whose $d_v(i,k)$ is less than T_d. In addition, the average value of $d_v(i,k)$ for these N_p^v elements is calculated as E_k^v, which is an index to measure the goodness of k to see whether it is a suitable translation. If E_k^v is sufficiently smaller and N_p^v is sufficiently larger, the position k is a good candidate of horizontal translation solution. Precisely, if $E_k^v \le T_e$ and $N_p^v \ge T_p$, k is collected as an element of the set S_x of possible horizontal translations, where T_p and T_e are two thresholds and set to 5 and 2, respectively. Through examining different k for all $|k| < W$, the set S_x can be obtained.

On the other hand, let P_a^h and P_b^h denote as the sets of horizontal edge positions in I_a and I_b, respectively. With P_a^h and P_b^h, we can define a distance function d_h as:

$$d_h(i,k) = \min_{1 \le j \le N_b^h} |P_a^h(i) - k - P_b^h(j)|, \quad (2)$$

where N_b^h is the number of elements in P_b^h. According to the distance d_h, with the similar method to obtain S_x, by examining different k for all $|k| < H$, the set S_y of possible vertical translations can be obtained. With S_x and S_y, the set S_{xy} of possible solutions can be obtained by: $S_{xy} = \{(x,y) \mid x \in S_x, y \in S_y\}$. Once S_{xy}

is obtained, we want to determine the best translation from S_{xy} through a correlation (block matching) technique. In this technique, the measure of sum of absolute intensity differences is used to measure the difference between two regions, i.e.,

$$D(p,q) = \sum_{x,y=-M}^{x,y=M} |I_a(x+p_x, y+p_y) - \mu_a - I_b(x+q_x, y+q_y) + \mu_b|, \quad (3)$$

where μ_i is the local mean of I_i and $(2M+1)^2$ the area of matching window. Due to the small size of S_{xy}, the best solution of translation can be obtained very quickly. More importantly, since our method uses edges to filter out all impossible solutions, more tolerance can be gained for overcoming different light changes. Details of the whole algorithm can be summarized as follows.

Edge-based Translation Estimation Algorithm:
I_a and I_b: two adjacent images to be stitched.

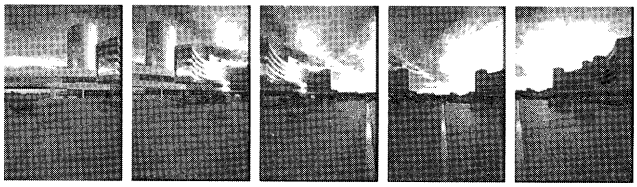
Step 1: Apply a vertical edge detector to find the sets P_a^v and P_b^v of vertical edge positions from I_a and I_b, respectively.
Step 2: Determine the set S_x of possible horizontal translations from P_a^v and P_b^v based on $d_v(i,k)$.
Step 3: Apply a horizontal edge detector to find the sets P_a^h and P_b^h of horizontal edge positions from I_a and I_b, respectively.
Step 4: Determine the set S_y of possible vertical translations from P_a^h and P_b^h based on $d_h(i,k)$.
Step 5: Let S_{xy} denote the set of possible translations, i.e., $S_{xy} = \{(x,y) \mid x \in S_x, y \in S_y\}$.
Step 6: Determine the best solution $T = (t_x, t_y)$ from S_{xy} through a block-matching technique.

In what follows, details of complexity analyses between the proposed algorithm and the normalized correlation technique are discussed. Assume that all input images are with the same size $N \times N$, a mask $M \times M$ is used to search desired translations, and the solutions range from $-R/2$ to $R/2$ in both the x and y directions. Generally, M is set to be a ratio of N, e.g., $M = N/5$. Then, the complexity of correlation techniques is $O(R^2N^2)$. In addition, the value of R is correlated with N, e.g., $R = N/4$. Then, the complexity of normalization correlation techniques will be $O(N^4)$. As to our proposed method, the complexity to extract different edge features is $O(N^2)$. With these features, a set of translation solutions will be extracted, i.e., S_{xy}, which includes $N_{S_{xy}}$ elements. Then, the complexity of the proposed method will become $O(N^2 + N_{S_{xy}}N^2)$. For general analysis, $N_{S_{xy}}$ is a constant or a ratio of R^2, e.g., $N_{S_{xy}} = \rho R^2$, where $\rho \ll 1$. Then, the complexity of our method becomes $O(N^2 + cN^2)$ or $O(N^2 + \rho N^4)$. Since $\rho \ll 1$, it is clear that our proposed method performs much better than other different correlation techniques in term of efficiency.

3. Experimental Results

In order to analyze the performance of the proposed method, four kinds of real images were used for testing. All the experiments were implemented on a personal computer with INTEL Pentium III 600 CPU. The size of image is 768×512 if a digital camera is used but becomes 320×240 if the image is extracted from a video sequence. The first experiment is a set of panoramic images. The second and third ones are the images with large intensity changes and moving objects, respectively. The last one demonstrates the success of mosaic construction and moving object detection with the proposed method.

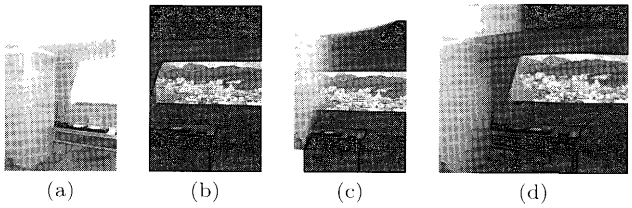
Figure 3 shows the case when the images are captured by a camera mounted on a level tripod. When these images are projected into a cylindrical model, only the translation parameters need to be estimated. Figure 3 (b) is the result got from the correlation technique. Due to trapping in local minima, the third and fourth images cannot be well stitched together. However, in Fig. 3 (c), all the input images are well stitched together with our proposed method. Figure 4 shows the case when images have larger intensity changes. (a) and (b) are the original images captured by a digital



(a)

(b) (c)

Fig. 3 Stitching result of a series of images captured by a rotated camera. (a) Series of panoramic images. (b) Result got from the correlation technique. (c) Result after stitching and blending.

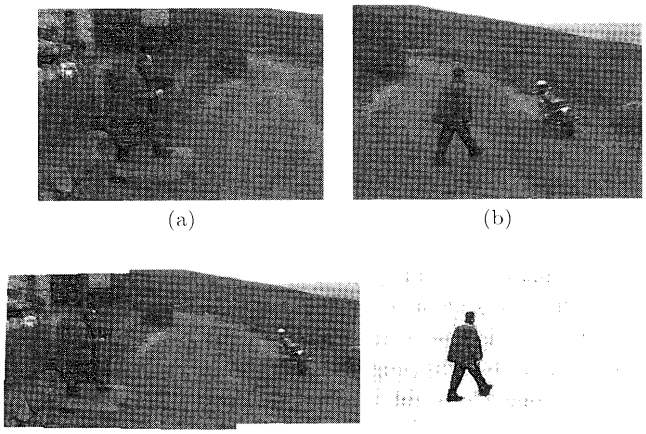


(a) (b) (c) (d)

Fig. 4 Stitching result when images have very large intensity differences. (c) Stitching result obtained by the correlation technique. (d) Result got from the proposed method.

(a) (b) (c)

Fig. 5 Stitching result when images have moving objects. (a) and (b) Images with a moving object. (c) Stitching result.



(a) (b)

(c) (d)

Fig. 6 Mosaic construction and object detection. (a) and (b) Input images. (c) Mosaic result. (d) Object detection result.

camera with the same size 768×512. The change of light conditions will lead to the instability of similarity calculation in techniques like block matching or phase correlation [4]. Therefore, in Fig. 4 (c), the correlation technique fails to register these images together. However, in our proposed method, edge features are first used to filter impossible solutions out. Therefore, our proposed method can own better capabilities to find desired correct parameters than the correlation techniques. Figure 4 (d) shows the correct result obtained by our proposed method. In above two experiments, the mask size for correlation calculation is set to be one sixth by one sixth of image dimension. Then, the proposed method used less than 0.1 seconds to find desired solutions, but more than 1.5 seconds were used for the correlation technique.

Figure 5 shows the case when images have some moving objects. The moving object will disturb the work of feature matching. However, our method still performs well to stitch these images together. Figure 6 shows that the proposed method also can be used to detect moving objects from video sequences. In Fig. 6 (a) and (b), the camera motion between them can be described by an 8-parameter affine transformation [2], [3]. With the proposed method, a good initialization can be provided to seek desired motion parameters through a simple optimization process. (c) is the stitching result and (d) is the detection result of moving object by image warping and differencing. The superiority of the proposed method can be verified through the preceding experimental results.

4. Conclusions

In this paper, we have presented an edge alignment method for estimating the translation displacements between images from sets of edge positions. Since edge positions are more robust than image intensity for overcoming light changes, the proposed method has better capability to resist different variations of image lighting. Different from traditional methods, the proposed method doesn't require building any correspondences or involve any optimization process. Therefore, the proposed method is very efficient and suitable for real-time applications and hardware implementation. Although the proposed method can handle only the translational camera movements, the estimation result can provide a good initialization for further optimization and analysis in various applications like video retrieval, video surveillance, and mosaic construction. Experimental results have shown our method is superior in terms of stitching accuracy, efficiency, and stability.

References

[1] H. Sawhney and S. Ayer, "Compact representation of video through dominant and multiple motion estimation," IEEE Trans. Pattern Anal. Mach. Intell., vol.18, no.8, pp.814–830, Aug. 1997.
[2] M. Irani and P. Anandan, "Video indexing based on mosaic representation," Proc. IEEE, vol.86, pp.905–921, May 1998.
[3] M. Bonnet, "Mosaic representation for video shot description," Proc. MPEG-7 Evaluation Ad Hoc Meeting, p.636, Feb. 1999.
[4] C. Kuglin and D. Hines, "The phase correlation image alignment method," Proc. IEEE Int. Con. on Cybernetics and Society, pp.163–165, 1975.
[5] S. Chen, "Quicktime VR-an image-based approach to virtual environment navigation," Proc. SIGGRAPH '95, pp.29–38, 1995.
[6] R. Szeliski, "Video mosaics for virtual environments," IEEE Computer Graph and Application, vol.16, pp.22–30, March 1996.
[7] H.Y. Shum and R. Szeliski, "Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment," Int. J. Comput. Vision, vol.36, no.2, pp.101–130, 2000.
[8] R. Szeliski and H.Y. Shum, "Creating full view panoramic image mosaics and environment maps," Proc. Computer Graphics Annu. Conf. Series, pp.251–259, 1997.
[9] I. Zoghlami, O. Faugera, and R. Deriche, "Using geometric corners to build a 2D mosaic from a set of images," Proc. Conf. Computer Vision and Pattern Recognition, pp.420–425, Puerto Rico, 1997.
[10] C.T. Hsu, T.H. Cheng, R.A. Beuker, and J.K. Horng, "Feature-based video mosaic," Proc. ICIP 2000, vol.2, pp.887–890, Vancouver, Canada, Sept. 2000.
[11] J.W. Hsieh, et al., "Image registration using a new edge-based approach," Computer Vision and Image Understanding, vol.67, pp.112–130, 1997.
[12] M. Sonka, V. Hlavac, and R. Boyle, Image Processing, Analysis and Machine Vision, Chapman & Hall, London, U.K., 1993.