A Panoramic Video System Based on Exposure Adjustment and Non-linear Fusion

Linlin Yang[1], Dandan Du[1], Baochang Zhang[1](✉), and Wankou Yang[2]

[1] School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China
{yanglin,bczhang}@buaa.edu.cn, ddnszbd@163.com
[2] School of Automation, Southeast University, Nanjing 210096, China
wkyang@seu.edu.cn

Abstract. This paper proposes a new video stitch method based on the exposure adjustment and nonlinear fusion. To solve the challenging problem of exposure difference between cameras, we propose the exposure adjustment method to deal with luminance difference among images in the YCrCb color space; As for the ghosting problem in the video stitch, we propose a nonlinear fusion algorithm based on, which achieves a much better performance than traditional linear fusion method, especially when there is a big disparity between cameras. The proposed method is real-time and efficient for a video surveillance system.

Keywords: Panoramic video · Image stitching · Image registration · Image fusion · SIFT

1 Introduction

Panorama, as an new technology, can expand view and display a wider range of scenarios at the same time [1]. The traditional video surveillance systems are generally using the fixed cameras with limited range of view, which can only monitor a fixed angle of space in front of the camera, thus fail to deal with all the events occurring within the range of around 360°. Currently, static panoramic image systems, except for a few applications using ultra-wide-angle lens or fisheye directly, mainly use software methods such as image stitching and fusion algorithms. In contrast, Dynamic video systems are largely rely on specific hardware: one solution is to use fast ball system, using a high-speed moving ball to capture a wide range of scenarios. But due to the limitation of rotational speed, it inevitably exists blind areas. Another approach is to use professional camera system equipped with panoramic cameras, but such systems are expensive, complicated to use and cannot be extended. In recent years, panoramic video surveillance system has aroused great attention [2][3]. Panoramic video surveillance system can simultaneously monitor all targets within the range of 360° at any time, and virtually eliminates monitoring blind spots and areas.

This paper focuses on the static monitoring system, and we briefly review the existing image stitching and fusion technologies [4][5], which can be used to show

J. Yang et al. (Eds.): CCBR 2015, LNCS 9428, pp. 499–507, 2015.
DOI: 10.1007/978-3-319-25417-3_59

the dynamic scene of a large-scale view in real time. The technology of stitching and fusion firstly conduct registration of a sequence of images which have overlapping regions between each other capturing by several cameras. And then those images are fused into a wide view mosaic image containing all the information of the image sequences, the so-called panoramic video. We do this from two images in the image registration process by extracting feature points in the overlapping region. Through matching the corresponding points, we can estimate image transformation relationship between them, and then do coordinates alignment according to the transformation. And after the boundary smooth transition of the jointing image, we form a confluent image containing the information of two images. The technology is different from the panoramic video technology using special instruments and PTZ video surveillance technology, which combines the advantages of both, thus it is of significant practical significance.

The rest of the paper is organized as follows, in section 2, we give a brief overview of our system, including the hardware components, the framework of our system and the specific algorithms we use for each step; in section 3, we briefly introduce the image stitching method used in our system, in section 4, we give some results to show the effectiveness of our system.

2 The Proposed Panoramic System

This paper designs a system which consists of several cameras with fixed installation in the relative positions, and the adjacent lens have certain overlapping scenes, so we can get multi-channel video streams at the same time. We discuss the process of the flow of camera video mosaic. The input is the four separate video cameras used to capture videos; the captured videos can enter the PC through the USB port for further processing, and then the wide-angle video will be displayed on the PC screen in real-time. Cameras used in the experiment are fixed on a tripod, the angle and direction of each camera can be adjusted within a certain range.

Fig. 1. The framework of camera video mosaic

The framework of our method is shown in Fig. 1, and each part will be descried in details as shown in the following subsections.

2.1 Image Preprocessing Based on Exposure Adjustment

In the process of the video acquisition and transmission, the image quality of the video is often affected by many unexpected factors. For example, the minor differences between two machines or the discrepancy in light, results in the difference among the several collected videos in terms of brightness and chromaticity; most of these noises are a random distribution. The existence of the above problems will not only affect the image matching precision in the process of image stitching, but also make the result of video stitching unsatisfactory. They can result in the discontinuity of the joint video images on both sides, and make the picture not clear enough to be seen. The image preprocessing is very necessary before jointing two images together to guarantee the quality of the video image stitching. The commonly used methods for preprocessing are histogram equalization, or other filter-based methods. In this paper, we propose to preprocess the image based on Exposure Adjustment method in the YCrCb color space.

Supposed there are two images, one is the standard image $img0$, another is to-be-adjusted image $img1$. Different from direct calculation of the mean of luminance Y for each image, we firstly divide the image into three blocks of different size, and then get the overall brightness mean $Iavg\_Y0$, $Iavg\_Y1$ in different weightings of each block and the standard deviation std_Y0, std_Y1. And then we calculate the average exposure $Savg\_Y0$, $Savg\_Y1$ of the two images according to formula(1). Y_curr is the current brightness value of $img1$, Y_new refers to the new brightness value. The methods are shown in following equations.

$$Savg\_Y = \frac{-(log^{\frac{255}{Iavg\_Y}-1.0})}{std\_Y} \,, \tag{1}$$

$$Y\_new = \frac{255}{1+exp(-std\_Y1 \times S\_curr)} \,, \tag{2}$$

$$S\_curr = \frac{-log^{\frac{255}{Y\_curr}-1}}{std\_Y1} + (avg\_Y0 - avg\_Y1) \times (1 + \frac{sign \times (Iavg\_Y1 - Y\_curr)}{255})$$

where

$$sign = \begin{cases} 1 & \text{if(Savg\_Y0} - \text{Savg\_Y1} \geq 0) \\ -1 & \text{otherwise} \end{cases} \,. \tag{3}$$

To illustrate the effectiveness of the proposed method, we show samples as in Fig. 2.

(a) (b)

Fig. 2. (a) The pictures before preprocessing (b) The pictures after preprocessing

It can be seen from the pictures, the exposure difference between two images decreased significantly after preprocessing. And thus we lay a good foundation for the following steps.

2.2 Image Registration

Image registration [6] is one of key steps of image mosaic. The current generic image registration methods can be divided into three categories: the image registration algorithm based on pixel values; based on transform domain, or feature-based image registration algorithm [7]. Considering that the geometric relationships between multiple cameras only need calibration once in the beginning step, we can adopt the methods with higher alignment accuracy. Among existing video mosaic methods, the SIFT algorithm based on multi-scale space theory is widely used. The SIFT method has a good robustness of translation, rotation, scale change, illumination change and so on [8], and enables our method handle images with varying orientation and zoom. Note that this is impossible to use traditional feature matching techniques such as correlation of image patches around Harris corners [9], Ordinary (translational) correlation. The reasons lie in that they are not invariant under rotation, or not invariant to changes in scale. Once features have been extracted from all images (number is n, with linear complexity), they must be matched. Since multiple images may overlap a single ray, each feature is matched to its k nearest neighbours (we use k = 2). This can be done in O(nlog n) time by using a k-d tree to nd approximate nearest neighbours [10]. If the ratio of the closest distance(d1) verse the second-closest distance(d2) is less than 0.4, we call that these matching is an inlier(correct matches) and preserve these matching. On the contrary, it will be called outlier (false matches) and discard these matches. The next step is to use the matched points to calculate the transformation matrix H which can warp from image I to image I'.

2.3 Robust Homography Estimation Using RANSAC

RANSAC(random sample consensus) [11]is essentially a sampling approach to estimating H. It is a robust estimation procedure that uses a minimal set of randomly sampled correspondences to estimate image transformation parameters, and nds a solution that has the best consensus with the data. In the case of panoramas we select sets of r = 4 feature correspondences and compute the homography H between them using the direct linear transformation(DLT) [12] method. We repeat this with n = 500 trials and select the solution that has the maximum number of inliers (whose projections are consistent with H within tolerance pixels). Given the probability that a feature match is correct between a pair of matching images (the inlier probability) is p_i, the probability of finding the correct transformation after n trials is

$$p(\text{H is correct}) = 1 - (1 - (p_i)^r)^n \ , \tag{4}$$

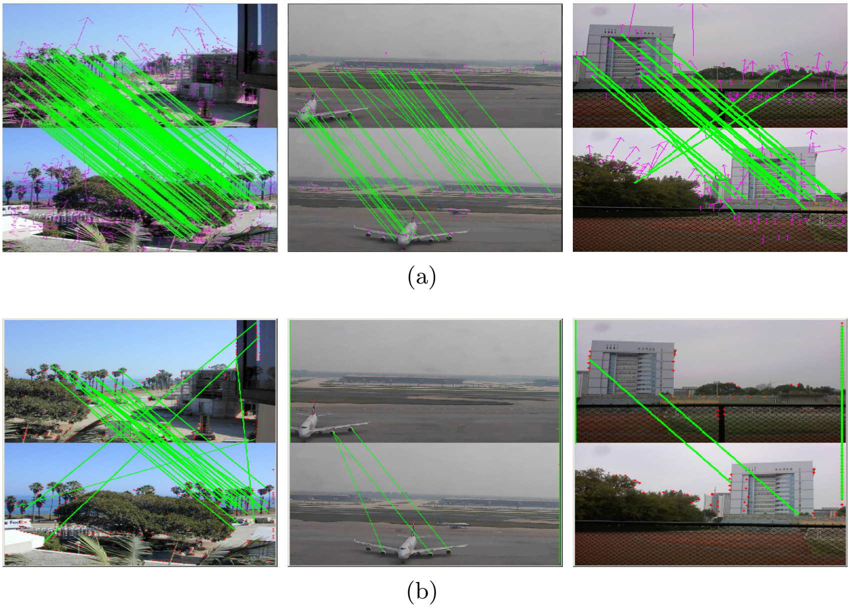


Fig. 3. (a) The results of SIFT feature detection (b) The results of Harris corner detection

then

$$n = \frac{\log(1-p)}{\log(1-(1-p_i^r))} . \tag{5}$$

The image matching results in Fig. 3 show that the SIFT algorithm can well adapt to various circumstances, despite a few false matching points, it has a good robustness of translation, rotation, scale change, illumination change and so on. Finally, after calculating a best-fit image transform from image feature correspondences using RANSAC, we finally get the perspective transformation matrix H. With the matrix H, we get the warp image.

2.4 A Non-linear Algorithm for Image Fusion

Images sampling in different time, under different light intensity could result in obvious seam in overlap on the stitched images. Fusion strategies should meet the requirements of two aspects: boundary transition should be smooth and can eliminate split seams to achieve seamless splicing; as far as possible to ensure no loss of original image information due to the split processing. Commonly used fusion algorithms include average fusion, linear fusion, multi-resolution fusion [13], etc. Our system adopts a kind of nonlinear fusion method, and the results show it can effectively eliminate the image ghosting caused by luminance difference and movement objects in the images.

$$I_{overlap}(x,y) = \alpha(x,y) \times I_1(x,y) + (1-\alpha(x,y)) \times I_2(x,y) . \tag{6}$$

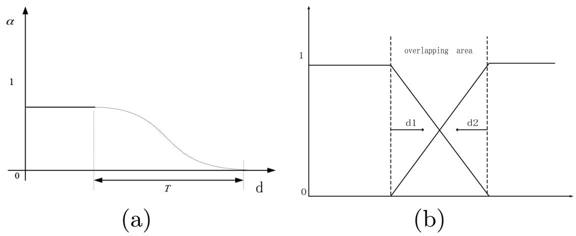


Fig. 4. (a) The changing curve of Non-linear fusion (b) The changing curve of Linear fusion

There are two cases for $\alpha(x, y)$,when we align the left image to the right,

$$\alpha(x,y) = \begin{cases} 1 & \text{if } \min(x, y, |x-W|, |y-H| > T) \\ \frac{\cos(\pi\cdot(\frac{\min(x,y,|x-W|,|y-H|)}{T}-0.5))+1}{2} & \text{otherwise} \end{cases}, \tag{7}$$

else, when we align the right to the left,

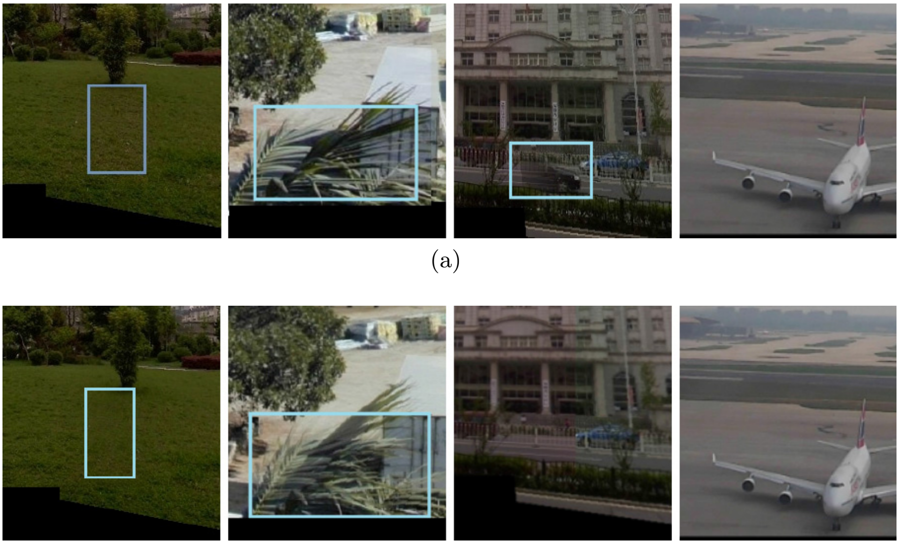
$$\alpha(x,y) = \begin{cases} 1 & \text{if } \min(x, y, |x-W|, |y-H| > T) \\ \frac{\sin(\pi\cdot(\frac{\min(x,y,|x-W|,|y-H|)}{T}-0.5))+1}{2} & \text{otherwise} \end{cases}, \tag{8}$$

where W and H represent the width and height of the original frame, T is the width of the nonlinear transition region, here it refers to the width of the overlapping area, as shown above. Since the non-overlapping area does not need to be weighted, we will resize the registrated weight templates according to the shape of the overlapping region, so that the weighting function only applies to the overlapping region. The value of a remains the same in the center of the frame, when getting closer to the boundary, namely into transition region of T, it will decline rapidly in a nonlinear form, and the decreasing rate can be controlled by the parameter T. We call this fusion method as nonlinear fusion. The comparison results of traditional linear method and our nonlinear method are as Fig. 5.

From the results under various situations as shown above, we can see that linear fusion algorithm contains ghost and fuzzy phenomenon in the overlapping region when there is large parallax between two images or moving objects, but the non-linear algorithm proposed in this paper, keeps the clarity of the main content of the scene, and greatly reduce the movement ghosting and fuzzy phenomenon.

3 The Image Stitching Method in Real-Time

The video surveillance system with multiple cameras needs some certain overlapping region between each camera. Under the circumstance of a few cameras, we usually use the method of frame-to-frame to calculate the transformation relationship between each two cameras and all the image sequences will be aligned



(a)

(b)

Fig. 5. (a) The results of Non-linear fusion (b) The results of Linear fusion

to one reference frame according to the transformations. But in the case of the system with large number of cameras, registering a large set of images introduces difficulties. In this paper, we adopt a method that combines the frame-to-mosaic and mosaic-to-mosaic methods to stitch video sequences of multiple cameras. Assume that there are some images C_i (i = 1,..., 4) captured by four cameras, we select one of the images (such as C_1) as a reference frame, and align C_2 to C_1, thereby generating a temporary stitching image M. Then align image C_3 to M, and update M using the C_3, and deal with C_4 the same way, in the end, we can get all the image alignment parameters.

In addition, an important element of a video surveillance system is to ensure its real-time performance, The SIFT features this paper choose has a good robustness of translation, rotation, scale change, illumination change, but the obvious disadvantage is its slow calculation speed, in order to overcome the shortcomings of slow calculation while maintaining precision of the SIFT features, this system learn to determine the homographic matrixes between each camera only using a few frames in the beginning, and once established, they will no longer change, then we project the video sequences onto global coordinate system using these alignment parameters, finally the small view video sequences can form a large view video sequence, so as to realize the real-time video processing.

Fig. 6. Real-time panoramic video of three camera

4 The Results and Conclusion

This paper designed and implemented a panoramic video surveillance system based on image stitching technology, the experimental results are shown in Fig. 6. By stitching and fusing the videos fixed in certain angle, we can finally get the output of 360° panoramic video images in real-time, and the generated panoramic images have high resolution, good visual effect and can guarantee high clarity of the scene. The light system is robust with high integration, easy to apply to financial systems, warehouses, prisons, mobile monitoring and many other occasions, especially suitable for indoor and outdoor monitoring system.

Acknowledgments. We acknowledged the support of the Natural Science Foundation of China, under Contracts 61272052 and 61473086, and the Program for New Century Excellent Talents of the University of Ministry of Education of China.

References

1. Brown, M., Lowe, D.G.: Recognizing panoramas. In: Proceedings of the 9th IEEE International Conference on Computer Vision, pp. 1218–1227. IEEE Press, Nice (2003)
2. Foote, J., Kimber, D.F.: Practical panoramic video and automatic camera control. In: Proceedings of the 1st IEEE International Conference on Multimedia and Expo., pp. 1419–1422. IEEE Press, New York (2000)
3. Rybski, P.E.: Camera assisted meeting event observer. In: Proceedings of the 21st IEEE International Conference on Robotics and Automation, pp. 1634–1639. IEEE Press, New Orleans (2004)
4. Xiangyang, Z., Limin, D.: An Automatic and Robust Image Mosaic Algorithm. Journal of Image and Graphics **9**, 417–422 (2004)
5. Dongmei, L., Yanjie, W.: Research of the Image Mosaic Method Based on Feature Point Match. Control and Automation **24**, 296–298 (2008)
6. Hsu, S., Sawhney, H.S., Kumar, R.: Automated Mosaics via Topology Inference. IEEE Computer Graphics and Applications **22**, 44–54 (2002)
7. Szeliski, R., Shum, H.Y.: Creating full view panoramic image mosaics and environment maps. In: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, pp. 251–258. ACM Press, Los Angeles (1997)

8. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision **60**, 91–110 (2004)
9. Harris, C.G., Stephens, M.: A combined corner and edge detector. In: Proceedings of the 4th Alvey Vision Conference, pp. 147–151. Elsevier Academic Press, Manchester (1988)
10. Beis, J.S., Lowe, D.G.: Shape indexing using approximate nearest-neighbor search in high-dimensional spaces. In: Proceedings of the 10th IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1000–1006. IEEE Press, Puerto Rico (1997)
11. Forsyth, D.A., Ponce, J.: Computer Vision: A Modern Approach. Pearson Education Limited, New York (2011)
12. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2004)
13. Burt, P.J., Adelson, E.H.: A Multi-Resolution Spline with Application to Image Mosaics. ACM Transactions on Graphics **2**, 217–236 (1983)