Real-time Moving Object Detection in Parallax Scene from PTZ Camera

Yushuang Zhang[1] ,Gaopan Huang[1],Hongxing Chang[1]
Integrated Information System Research Center
[1]Institute of Automation, Chinese Academy of Sciences
Beijing, P.R. China

Abstract-This paper presents a real-time moving object detection algorithm from a video sequence taken by a PTZ camera. In order to extract the moving foreground, previous approaches almost are based on static background. For a moving camera, many solutions using mosaic background subtraction have been proposed, which offer real time capabilities or high quality of the detected objects. However, most of them rely on prior information or the algorithms work with a depth field of view only. In this paper we propose a novel solution to achieve an automatic and high quality mosaic background of a parallax scene and apply it to moving object detection based on background subtraction algorithm. Accurate experiments performed in outdoor sequences assess the quality of the mosaic as well as of the detected moving masks.

Keywords-image mosaic; moving object detection; histogram matching; SURF descriptor

I. INTRODUCTION

Moving object detection plays an important role in video surveillance system. In the last few years, many solutions have been proposed, such as *background modeling*, *background subtraction*, *optical flow* and so on, to detect moving objects. The *background subtraction* technique is known as the one yielding the highest quality of the detected moving masks when using conventional stationary cameras. However few algorithms have been proposed in literature to perform mosaic background detection with Pan-Tilt-Zoom(PTZ) cameras. Most of them need to exploit prior information about scene or camera settings in order to achieve an accurate motion estimation meanwhile fulfilling the real-time requirements [1, 2]. Actually, it is often difficult to extract prior information from the hardware utilized, and also unreliable for some sensors working longtime or camera calibration inaccuracy. One of the heaviest drawbacks of background subtraction algorithm for PTZ cameras is the computational burden needed to achieve high quality mosaics in real time.

Therefore, sometime the approaches used to achieve a good mosaic forces the authors to perform a background subtraction offline [3], or limit the application fields characterized by a depth field of view or aerial view without parallax [4].

In this work we present a real time, coarse-to-fine positioning, moving object detection algorithm based on mosaic background of the parallax scene. It is divided into two parts as shown in Fig.1.Firstly,we get the background panorama image by estimating the transform relationship based

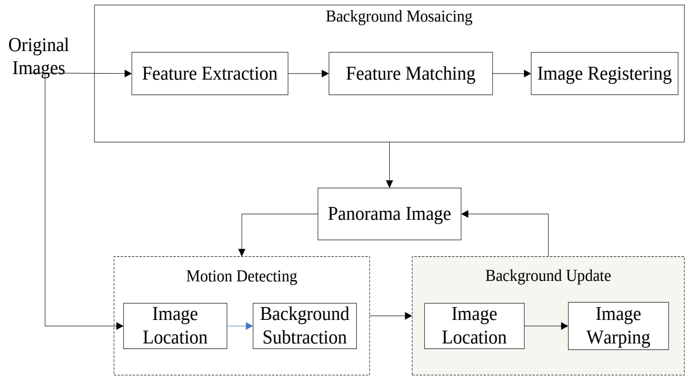


Figure. 1. Block diagram of the proposed algorithm

on the keypoint matching approach between successive images with small overlap and arbitrary rotation around the optical axis of the PTZ camera. Then, accurately locate the new frame to the mosaic panorama image to detect moving object through background subtraction algorithm or update background when necessary.

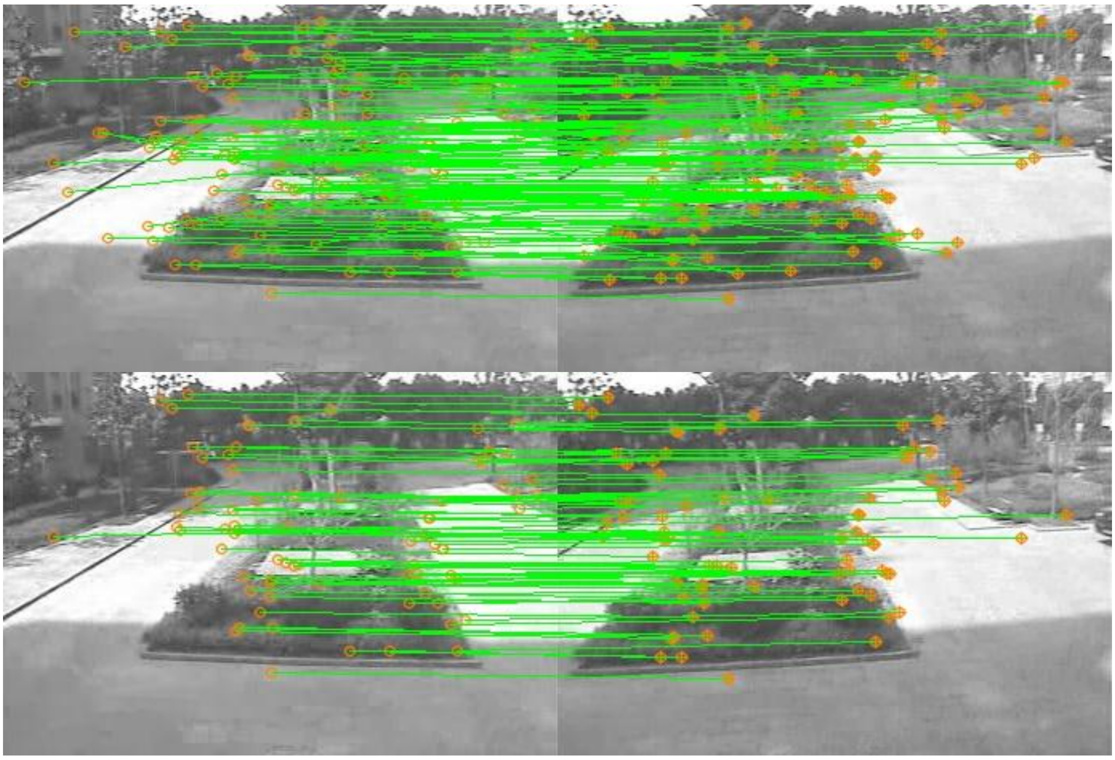
The paper is organized on the basis of the structure of our algorithm. In Sections 2 we presents the background mosaic algorithm. In Section 3 motion detection is described. Experimental results are shown in Section 4. At last, a conclusion of this paper in Section 5.

II. BACKGROUND MOSAICING

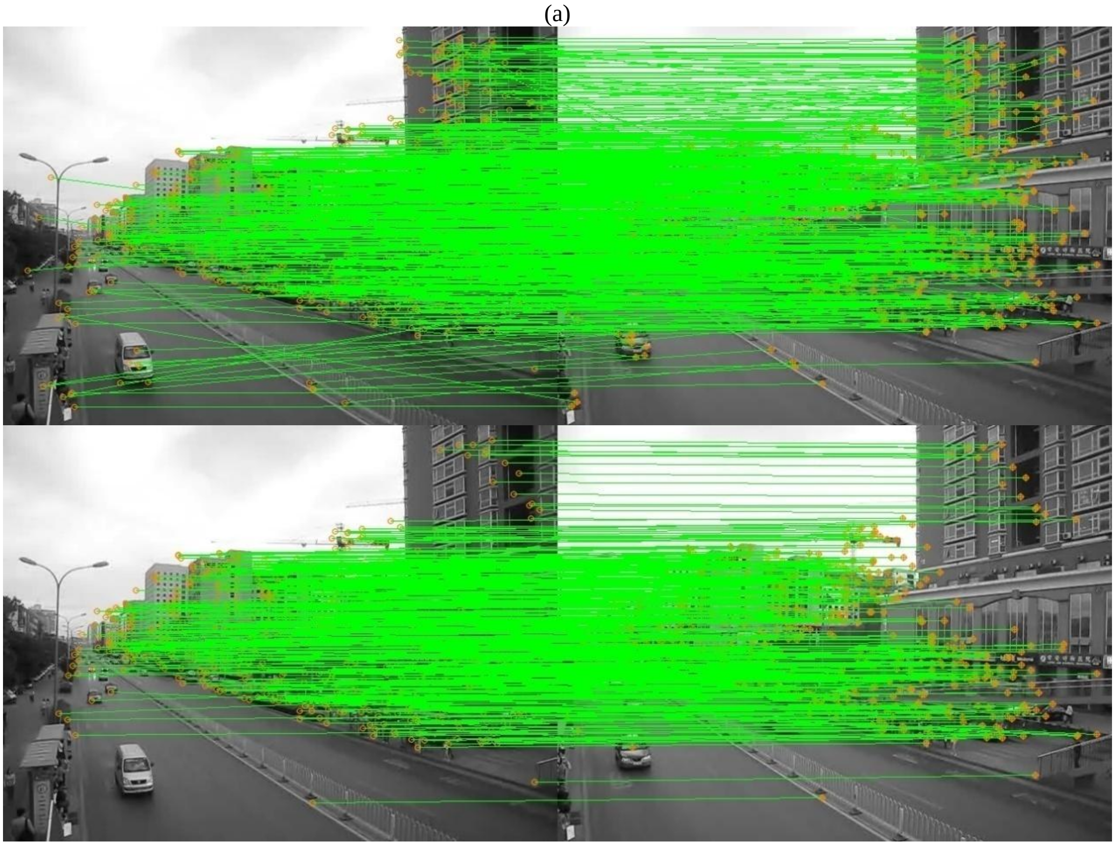
Many methods are known to perform spatial image alignment, based on phase correlation, pixel intensities SSD minimization or feature tracking. The first ones are mostly used to build up mosaic images by collecting frames coming from rigid or, at most, affine camera motion. The second ones bear a heavy computing burden for real time applications. In order to meet the real time requirements when dealing with projective transforms, we choose the feature-based approach.

For each frame, a set of feature points are extracted and matched with the ones from consecutive frame. The matched feature points are utilized to estimate the camera's ego-motion. For the parallax scene, the transformation between two consecutive images acquired by a moving camera can be regard as projective transformation.

The matched points determine the model's parameters for each couple of sequence frames, and permit to merge images into one seamless picture. Nevertheless, the mismatched points, which always appear in image registration because of some noise or indistinguishable points, bring in inescapable errors, especially when the number of original keypoints are not many. Therefore, eliminating outliers is important before estimating



(a)



(b)

Figure 2. Matched feature points of two images from rotating camera(pan only).(a)and(b) images have different qualities and depth. The top row shows feature points matched by SURF descriptors, and there are some mismatched point pairs (i.e. some slashes in the images).

the parameters. Epipolar constraint [5,6] and a simple yet efficient clustering method [7] is used to filter out corner points which exhibit inconsistent motions (e.g. errors in feature matching or features belonging to moving objects).In [8],the Median Flow Filter method is fast and does not need any specific knowledge of the scene or motion, which is efficiency for refining matches. P. Smith calculated the median flow by taking the mean of the most tightly bunched *n* vectors from the *k* nearest vectors. For each match, he tested the direction first and then their length. But in our work, most of the matches are correct and the values of their lengths and angles are similar, and only adopting the median value of the middle *n* vectors is enough. For each match, if the point in a similar direction or a similar length when compared to the median flow, it is classified as an inlier, otherwise it is discarded as an outlier.

Fig.2 shows matched pairs of feature points before (up) and after (down) eliminating mismatched points. Using our approach, most of wrong pairs are eliminated. In Table 1, we list the related results to show the high rate of inlers by using our approach.

After refining keypoints, the homography parameters are almost exact. To mosaic a sequence of frames into a panorama image, some approaches have been researched. One is the well known *frame to frame* (or pairwise) registration. The efficiency of the method is paid in terms of registration locality. That is, each couple of frames is registered without considering previous registrations, thus preventing all sequence frames from being global coherent. The other is global registration approach. It makes up for this limitation with the possibility of aligning groups of frames by considering both their spatial and temporal contiguity. Of course, the drawback of this approach is the need of the whole sequence to be known in advance, thus making global methods suitable mainly for batch computation. At last, the *frame to mosaic* technique performs a sort of "back registration" between the current frame and the existing mosaic. This method represents a trade between *frame to frame* and global registration.

In our work, to mosaic a 360-Degree panorama image, the *frame to mosaic* method is utilized. For the real time background mosaic, we do not know which frame should be classified as background image in advanced. For two frames captured from different view angles, the size of overlap area decides the number of matched feature points. The bigger the size is, the bigger the number is. So if the frame has small overlap area with the previous background mosaic image, it should be mosaic to the background. We use a threshold to choose frames for the background mosaicing. The value of the threshold is very important. If it is too smaller, the distance of the two consecutive frames selected as background images is so big that probably bring in errors during image registration. Else if it is too smaller, the scene of a new frame, which has been mosaic to the background image, may be added into again. For the video composed with cycle repeated scene when the PTZ camera working as a monitoring mode, when capturing a new frame, we extract keypoints first and match them with those ones extracted from the previous frame, and then it will be stitched with the mosaic image if the number of matched keypoints is smaller than the threshold , otherwise, the work of mosaic background image is completed. From then on, we detect moving objects or update background when inputting new frames.

There is a big problem in building background mosaic of a large scene. If mosaicing many frames to one image, it has incremental errors and big shape distortion because of cumulative multiplication of the homography matrixes and large change of view angles. To avoid this problem, a global registration technique bundle adjustment [9] can minimize the registration errors among all the frames and get an optimal mosaic result, but it requires the entire sequence to be known in advance. In our work, we use a serial of mosaic images instead of one panorama image, each of which ranges about 60 degrees, and all of them make up the 360-Degree panorama image. Meanwhile, we let the last frame of the mosaic image be the first frame of the next one, which efficiently prevents remaining gaps between the consecutive mosaic images.

III. MOTION DETECTION

Having a reliable background panorama image, we can directly extend the use of background subtraction algorithm for a stationary camera to a moving PTZ camera. The main issue in this case is to find out which part of the mosaic background the current frame corresponds to and the transformation relation-ship between the current frame and the region it corresponds to.

In our work, it is a coarse-to-fine scheme, rough and fine positioning, to locate current frame to the corresponding region in a serial of mosaic images, which first find out which one mosaic image and the rough region the current frame corresponds to and then precisely locate its coordinates.

A. Rough Positioning

Histogram matching is often applied to the image processing and computer vision field, such as i*mage registration*, *object detection* and *tracking*, because of its high efficiency and accuracy. By comparing two different images' histogram distributions, we can judge whether they are from the same scene. In our system, we need to locate the current frame in a serial of mosaic images to find its position using histogram matching. Some classic similarity measurement methods such as *Bhattacharyya* distance and *cross-correlation*, can be used for comparing two histograms .

Since the full search algorithm is complexity and time-assuming, we use an optimization search approach to speed up it. If *k* represents the ratio of the translation amount to image width, when moving the window along x-ordinate, and *d* is the similarity of the two histograms from consecutive overlapping blocks. We select 9 frames from the video sequence ranged from frame 1 to frame 150. Fig.4 shows the relation between *k* and *d* of these frames. Here *k* ranges from 0.1 to 0.5. We can see that when k is varied over a range from 0 to 0.3, *d* decreases linear with the increasing of *k*. In order to ensure the universality of the result, we test it in more than 100 frames, and the the result remains exactly the same.

So d can be approxi-mately represented by (1):

$$\begin{cases} d = r_1 * k + a_1 & 0 < k < 0.2 \\ d = r_2 * k + a_2 & 0.2 < k < 0.3 \end{cases} \quad \textbf{(1)}$$

Where $r_1$and r_2 are negative,a_1 and a_2 are approximately 1.Search step k and similarity d meet piecewise linear relationship in the qualifying range from 0 to 0.3.Roughly, when k ranges from 0.1 to 0.9, it can be seen as a linear relationship.

In Fig.5, 20 different frames are tested to study the change of the similarity d linked to the step k. It shows that only when k=0.1, d is always greater than 0.9.In other cases, d varies greatly. This result remains the same in much more frames being tested. So using these conclusions above, we set a search principle by setting a threshold T_d for d. The value of k is determined as follows:

$$\begin{cases} k = 0.1 & d \geq T_d \\ k = 0.2 \sim 0.3 & d < T_d \end{cases} \quad \textbf{(2)}$$

In the region close to the optimal location, the search step k should be smaller, otherwise greater. Let k be 0.1 to keep the similarity more than 90% between current and the next searching region when $d \geqslant T_d$. Or be the value ranged from 0.2 to 0.3 to enlarge search step and guarantee d linear decrease.

The fast search algorithm will be achieved by utilizing the algorithm discussed above. Namely, after checking a candidate block, the value of k will be set according to (2) .This variable step search algorithm can generate almost the same accuracy as that generated by full search.

B. Fine positioning

After rough positioning, we get rough ordinations of the current frame corresponding to the mosaic images. In this step, we accurately locate the ordinations of the frame in the background mosaic images. Feature point matching is an important algorithm for image registration, and matching by descriptors has high qualities of speed and accuracy. In [10], K.Mikolajczyk analysed some main feature point descriptors, such as GLOH, gradient location, SIFT and so on. Comparing with other descriptors, SIFT has the qualities of good discernibility, rotation invariant and insensitive to illumination changes. SURF [11], improved on SIFT, has similarly qualities but a higher speed.

In our system, we use SURF detector and descriptor to extract and match feature points respectively. Since the rough position has laid out the corresponding region, we just need to extract feature points in a small part of the full mosaic image.

IV. EXPERIMENTAL RESULTS

Extensive experiments have been accomplished in order to assess the quality of background mosaic and of the moving objects achieved by the background subtraction. As the first figures, we show sequences of mosaic images as background model. The Sequence S1,as shown in Fig.5, has been acquired at 10fps by using a PTZ camera fixed on a head and pure panning moving. In Sequence S1, each background image is made up of four frames, and the interval of each frame is 2s. All of these mosaic images make up the panorama background image. In these mosaic images you can see the common regions of every two frames joint very well but a marked seam can be seen. This is caused by illumination change in different frames. For a good vision effect, we can use image fusion methods to get a seamless mosaic image. But it is not necessary for motion detection in our work. The static cars are considered as background in these frames captures from a short video. In the future research, we can update the background images using more frames. The results of our moving object detection referring to the same environments depicted in Fig.5 are shown in Fig.6, where the detected moving masks have been superimposed on the real moving objects. In Fig.6 (a) a moving car is detected, and plants and buildings are not detected as moving objects in the scene, thus confirming the ability of separating motion coming from foreground objects and the camera itself. Fig.6 (b) a moving person is detected. Here the speed of camera panning is higher than person's. Although the foreground object changes much less than the background scene, the moving mask always adheres to the moving person perfectly and no alignment errors are introduced by such a movement.

V. CONCLUSION

In this work we have presented a real-time system to detect moving objects from a video sequence taken by a PTZ camera, through the background subtraction technique. Many original solutions have been adopted and improved in order to meet real time requirements and the high quality of the mosaic needed to achieve extremely reliable masks of the moving objects. These effective and efficient algorithms allow us to distinguish between camera and foreground object motion, also yield excellent results both in terms of mosaics and moving object detection.

Our current system still has some limitations. First, if the background scenes do not have sufficient features, the mosaic image will contain large errors, and the motion detection may not work well. Second, when the grayscale of the foreground and that of background are very similar, it also has difficulties to distinguish them.

REFERENCES

[1] S. Kang, J. Paik, A. Koschan, B. Abidi, and M.A. Abidi, Real-time video tracking using PTZ cameras. Proc. of the 6th International Conference on Quality Control by Artificial Vision (QCAV'03), Gatlinburg, TN, USA, Vol.5132,pp.103–111, May 2003.

[2] K.S. Bhat, M. Saptharishi, and P.K. Khosla. Motion detection and segmentation using image mosaics. Proc. of IEEE International Conference on Multimedia and Expo, Vol.3,pp.1577–1580, July 2000.

[3] A. Bartoli, N. Dalal, B. Bose, and R. Horaud. From video sequences to motion panoramas. Proc of the Workshop on Motion and Video Computing (MOTION'02), pp.201-207, December 2002.

[4] Y. Sugaya and K. Kanatani. Extracting moving objects from a moving camera video sequence. Proc. of the 10th Symposium on Sensing via Image Information(SSII2004), Yokohama,Japan, pp.279–284, June 2004.

[5] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. Cambridge University Press, 2004.

[6] Chang Yuan, Ge ´rard Medioni, Jinman Kang, Jinman Kang. Detecting Motion Regions in the Presence of a Strong Parallax from a Moving Camera by Multiview Geometric Constraints. IEEE TRANSACTIONS

ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 29, NO. 9, SEPTEMBER 2007

[7] A. Bevilacqua, L. Di Stefano, and P. Azzari. An effective real-time mosaicing algorithm apt to detect motion through background subtraction using a ptz camera.In Proc. of AVSS, volume 1, pages 511–516, 2005

[8] P. Smith, D. Sinclair, R. Cipolla, and K. Wood, "Effective. Corner Matching"in Proceedings of the 9th British. Machine Vision Conference, pp. 545–556, 1998M

[9] Triggs B,et.al. Bundle adjustment-a modern synthesis. In International Workshop on Vision Algorithms, Springer, Keykyra, Greece, 1999

[10] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," Proc. Conf. Computer Vision and Pattern Recognition, pp. 257-264, 2003

[11] Herbert Bay,Tinne Tuytelaars, and Luc Van Gool.SURF:Speeded Up Robust Features. Computer Vision–ECCV 2006,Springer.

TABLE I. INLIER RATES BEFORE AND AFTER FILTERING

	Fig. 2 (a)	**Fig. 2 (b)**
rough matched pairs	145	780
the number of inliers	121	651
inliers rate(%)	83.45	83.46
matched pairs after filtering	77	560
the number of inliers after filtering	71	521
inlier rate(%)	92.21	93.04

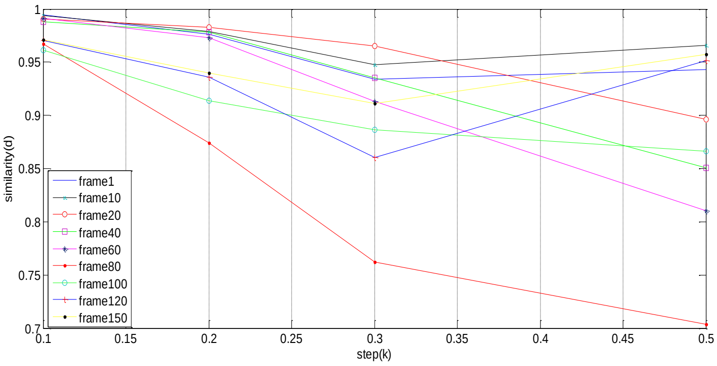


Figure 3 . Relation between *k* and *d*

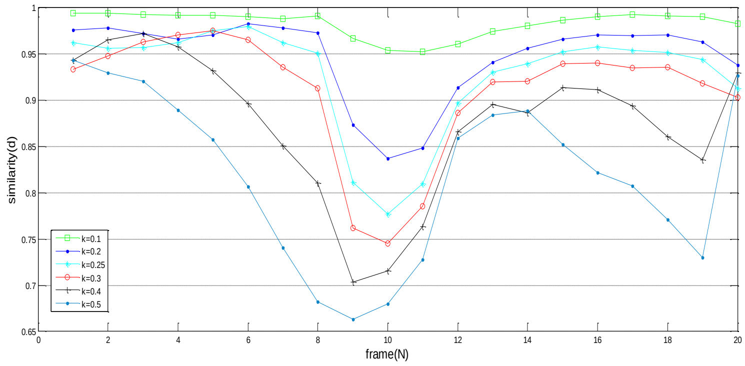


Figure 4. Similarity rates *d* for the step *k* change.

Sequence S1. outdoor ,pure panning.
Figure 5.Background mosaic images

(a) Detect a moving car in S1 .

(b) Detect a moving person in S1.

Figure 6. Moving objects detection in Sequence S1.