Representation of the player action in sport videos

Jingwen Zhang* Jiankang Qiu* Xiangdong Wang† Lifang Wu*
*College of Electronic Information and Control Engineering, Beijing University of Technology, Beijing, China,
E-mail: zhangjingwen@emails.bjut.edu.cn, qiujiankang1@emails.bjut.edu.cn
†China Institute of sports science, Beijing, China
E-mail: wangxiangdong@siss.cn

***Abstract*—In a sport video, the complex and dynamic background and the complexity of player actions bring considerable difficulty to analyze the player's action. In this paper we propose a method to represent the player's action on the panoramic background in a whole. It is helpful for watching and analyzing the player's action. We first build the panoramic background by the inter-frame relationship. Then object segmentation is implemented by subtraction between the panoramic background and each warped frame. Finally, the player's action in each frame is represented on the panoramic background. And the trajectory of the player action is presented on the background. The experimental results show that the proposed method is effective for the sports video like diving, jumps, which the player performs his or her action in a large arena, and the camera motion mainly includes horizontal and vertical direction.**

I. INTRODUCTION

With the explosive growth of digital videos in our daily life, automatic video analysis has become a basic requirement. In recent years, the analysis of sports videos has attracted great attention due to its mass appeal and commercial potentials. Many researches have been developed. For example, Wang et al. [1] and Xie et al.[2] proposed an approach to analyzing the structure of sports video automatically. Yu and Jiang et al. [3] presented a novel algorithm to automatically acquire accurate camera calibration from broadcast tennis video. The obtained camera matrix is utilized in 3D virtual content insertion. It was also utilized for tennis-ball detection and tracking. Wu et al [4] presented a progressive algorithm to locate the balls on the table and estimated their colors.

In some individual sports videos (such as diving, jumps, gymnastics videos, and so on , as shown in Fig. 1), the player performs his or her action in a large arena and the camera needs to be operated with pan/tilt/zoom to capture the player in the middle of the image. Unlike the general watchers, the coachers or the players pay more attention to the accuracy of the action of the players. Frame by frame analysis is important. The manual analysis is labor-intensive and time-consuming. Therefore, it is important to develop an automatic system of sport video analysis.

Many works focus on this kind of videos analysis. In [5], based on a coarse separation of foreground from images and the dynamic background construction technique, S. Wu et al constructed an accurate background for current frame, extracting the contour of the object by subtraction. In [6], F. Odone et al estimated the dominant motion of the scene by tracking features over the video. The moving objects get removed in the final mosaic by computing the median of the grey levels. A. Colombari et al. [7] achieved segmenting and connecting one or more foreground moving objects. H Li and J Tang et al. [8] automatically segmented the highlights in the video clips and recognized the action types to support action-based video indexing and retrieval.

In this paper, we present an integrated coarse-to-fine approach to analyzing the sports video and to presenting the player's action using the trajectory of the action. First, we estimate the global motion by using the stitching method. The dominant motion is obtained by calculating the homographies between video frames with RANSAC technique that removes the outliers. A global 2D motion model of the whole image is obtained using the reliable motion information of feature points. We warp each image to the common plane based on the calculated homographies and only use the overlapping method during blending step to obtain the final panoramic background. The experimental results using diving video show the efficiency of the proposed approach.

The remaining parts of the paper are as follows: In Section 2, we describe the global motion estimation and background construction using stitching method in details. We describe how the transformation between a pair of images is computed. In Section 3, we present the object segmentation algorithm and the action representation in the panoramic background. Section4 is the experiment results. Finally the paper is concluded in Section 5.

Fig. 1. Some sports actions

II. CONSTRUCTION OF THE PANORAMIC BACKGROUND

In this paper, we want to present the player's action using the trajectory of the action, so global estimation becomes a basic technology for this dynamic background sport video. We first estimate the inter-frame mapping based on feature matching. Then we align all the frames to the initial frame by the inter-frame mapping relationship. Finally, the panoramic background is obtained by fusing the background regions in all the frames.

2.1 Estimation of the inter-frame transformation

In this paper, we adopt the 6-parameter affine model [9] to describe the camera motion because it is computationally efficient and powerful to model the translation, rotation, and scaling operations of camera when the relative depth of the object is not large, which is applicable to most videos including our case.

Let m =(u, v) be the position of a point P in the current frame I_k , and $\mathbf{m}'=(u',v')$ be the position of point P in frame I_{k+1}. We define the 3*3matrix H as in Equation (1) is the affine model between the frames. It has six degrees of freedom and needs at least three pairs of points to get all parameters. So we can build the relation between I_k and I_{k+1} using the affine model as in Equation (2).

$$H=\begin{bmatrix} H_0 & H_1 & H_2 \\ H_3 & H_4 & H_5 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} u' \\ v' \end{bmatrix}=\begin{bmatrix} H_0 & H_1 \\ H_3 & H_4 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix}+\begin{bmatrix} H_2 \\ H_5 \end{bmatrix} \tag{2}$$

SIFT features [10] are located at scale-space maxima/ minima of a difference of Gaussian function. They are invariant under rotation and scale changes, giving some robustness to affine change. Then the k-d tree is used to find the approximate nearest points. Next, the RANSAC (random sample consensus) [11] algorithm is utilized to find the transformation matrix between the two adjacent images. It continuously samples a minimal set of corresponding points, repeating to estimate image transformation parameters. Then it will find the best homography matrix that has the maximum number of inliers. Finally, we remove the outliers based on this matrix.

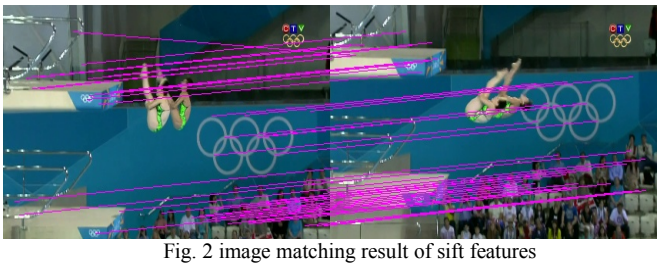


Fig. 2 image matching result of sift features

(a) (b)

Fig.3 (a) the warping image based on transformation matrix (b) misacting result of the two images in Fig (2)

2.2 Construction of the panoramic background

We use the initial frame as the reference frame. Based on affine matrix between frames, the panoramic background will be generated by warping the image to the reference plane frame by frame.

From the inter-frame transformation matrixes, we could align all the frames to the reference frame. Let's assume, we have total N frames in the sports video. So that we could obtain the N-1 pair inter-frame transformation matrix, they can be represented as $H_1, H_2, H_3, H_4 \cdots\cdots\cdots H_{N-1}$ .For the *i*th frame, we could obtain the transformation matrix between it and the reference frame:

$$RH_i = H_1 * H_2 * H_3 \bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet H_i \tag{3}$$

After Obtain the transformation matrix H' to the reference frame, we do warping for each image. The traditional blending method, for example, adding and the averaging method, will lead to the superposition because of the presence of the moving foreground. So here we use a very simple approach: we make the reference frame direct overlay the warping image, so the moving foreground seems "disappearing". As shown in Fig.3, it is the result of the reference frame directly overlaying the warping image (a). We do the same work to other frames and the final background is constructed as shown in Figure 4.

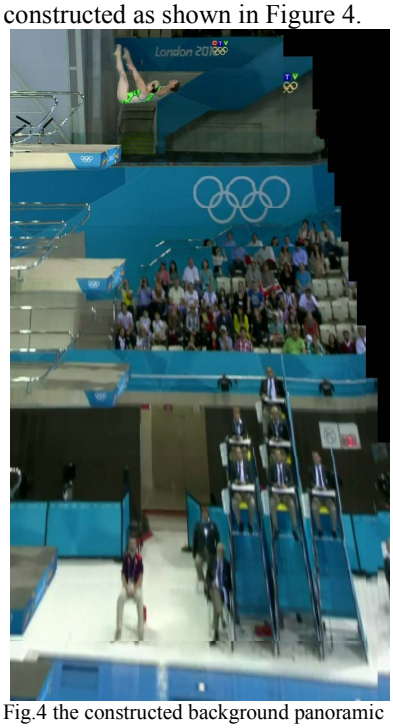


Fig.4 the constructed background panoramic

III. OBJECT SEGMENTATION AND ACTION REPRESENTATION IN THE BACKGROUND PANORAMIC

By now, we get the panoramic background and the warped image of each frame. We can easily segment the moving foreground by background subtraction. The binary difference image is obtained by thresholding. The threshold is determined by the experiments. The example result is shown in Figure 6.

There is a lot a noise after the subtraction. In order to get the final complete mask of foreground, the post processing is needed. We first filter out the noise by erosion. Then we dilation is implemented two times, the result is shown in Figure 7. At last, we obtain a complete moving foreground region by area constraints of the max area region to remove other noise.

After getting the foreground mask, we can paste the moving object back to the panorama and obtain the final effect picture, as shown in Figure 8(a).

For each region its center in the panorama background is computed, the trajectory of the player action could be obtained by connecting the center of the moving object, as shown in Figure 8(b).

(a) (b)

Fig.5 The mapping relationship between the frame and the panoramic background. (a)the panoramic background (b)the mapping relationship between the frame and the panoramic background

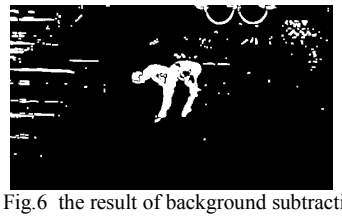
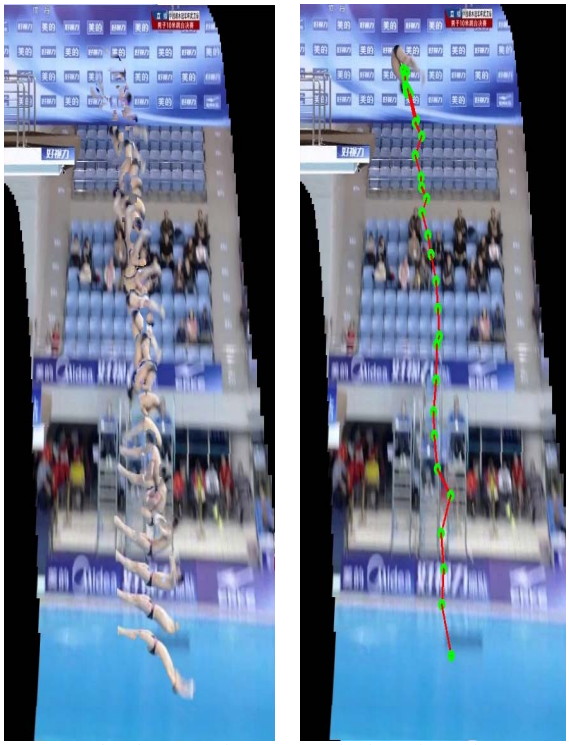


Fig.6 the result of background subtraction

Fig. 7 the output of post processing from Figure 5



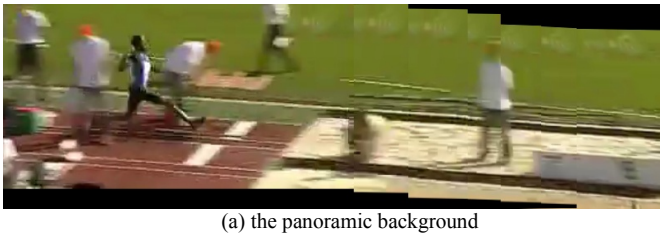
(a) the player's action (b) the action trajectory

Fig.8 the player's action and the trajectory of action in the panoramic background

IV. EXPERIMENTAL RESULTS

We test the proposed approach using two videos: a diving, a long jump video and a skiing video. Our approach gives the correct results successfully for diving video and long jump video. But it fails in skiing video. The diving video's results are shown in Figure 8. The long jump video's results are shown in Figure 9. These results show that our approach is effective to both horizontal and vertical camera motion.

We further analyze why our approach fails in skiing video. We find that construction of panoramic background is based on the inter-frame transformation matrix， which is computed from the pairs of matching points. In the skiing video, the background is too simple, and it is difficult to find the matching pairs between the inter frames. Therefore, we could not construct the panoramic background from the skiing video.



(a) the panoramic background

(b) the player's action

Fig. 9 The results of a long jump video

V. CONCLUSION

In this paper we propose a method to represent the palyer's action. We first construct the panoramic background; object segmentation is implemented by subtraction between the panoramic background and each warped frame. The experimental results show that the proposed method is effective for the sports video like diving, jumps, which the player performs his or her action in a large arena, and the camera motion mainly includes horizontal and vertical direction.

REFERENCES

[1] F. Wang, J.-T. Li, Y.-D. Zhang, and S.-X. Lin, "Semantic and structural analysis of TV diving programs," *J. Comput. Sci. Technol.*, vol. 19, no. 6, pp. 928–935, Nov. 2004.

[2] L. X. Xie, P. Xu, S.-F. Chang, A. Divakaran, and H. Sun, "Structure analysis of soccer video with domain knowledge and hidden Markov models," *Pattern Recognit. Lett.*, vol. 25, no. 7, pp. 767–775, May 2004.

[3] X. G. Yu, N. J. Jiang, L.-F. Cheong, H. W. Leong, and X. Yan, "Automatic camera calibration of broadcast tennis video with applications to 3-D virtual content insertion and ball detection and tracking," Compute. Vision Image Understand., vol. 113, no. 5, pp. 643–652, May 2009.

[4] L.F Wu, J. Liu, Z. H Cheng, "An effective multi-object detection approach". International Symposium on Intelligent Signal Processing and Communication Systems. 2010:1-4

[5] S. Wu, S.-X. Lin, and Y.-D. Zhang, "Automatic segmentation of moving objects in video sequences based on dynamic background construction," Chin. J. Comput., vol. 28, no. 8, pp. 1386–1392, Aug. 2005.

[6] F. Odone, A. Fusiello , E. Trucco "Layered Representation of a Video Shot with Mosaicing" Pattern Analysis & Applications (2002)5:296–305

[7] A. Colombari, A. Fusiello, V. Murino "Segmentation and Tracking of Multiple Video Objects" Pattern Recognition, 2007

[8] H Li, J Tang, S Wu, Y Zhang "Automatic Detection and Analysis of Player Action in Moving Background Sports Video Sequences" IEEE transactions on circuits and systems for video technology , Vol. 20, No. 3, March 2010

[9] M Brown, DG Lowe "Automatic Panoramic Image Stitching using Invariant Features" International Journal of Computer Vision, 2007

[10] DG Lowe "Distinctive image features from scale-invariant keypoints" International journal of computer vision, 2004

[11] M. Fischler and R. Bolles. "Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography". *Communications of the ACM*, 24:381–395, 1981.