Automatic Video-based Analysis of Athlete Action

Haojie Li[1,2], Shouxun Lin[1], Yongdong Zhang[1], Kun Tao[1,2]
[1]*Institute of Computing Technology, Chinese Academy of Sciences, 100080, Beijing, China*
[2] *Graduate University of the Chinese Academy of Sciences, 100080, Beijing, China*
{hjli, sxlin, zhyd, ktao}@ict.ac.cn

Abstract

Video analysis of athlete action is becoming an important tool for sports training, since it has no intervention to the athlete and there are abundant archived data it can exploit. In this paper we present our work on the automatic analysis of complex individual actions in diving video aiming at providing biometric measurements and visual tools for coaching assistant and performance improving. The main body joint angles of the athlete are automatic obtained by 2D articulated human body model fitting and shape analysis techniques. Two visual analyzing tools: motion panorama and overlay composition, which are extremely suitable for individual sports game training are presented. The encouraging experimental results show the effectiveness of the proposed system.

1. Introduction

Recent advances in video technology and computer processing performance have motivated the research of using computer vision and image processing techniques to analyze sports game video for tactics statistics, computer-aided coaching or performance improvements. Compared with previous methods that require retro-reflective markers or magnetic sensors to be placed on an athlete body, video-based approach has the advantages that it is much less expensive and does not interfere the performance of athlete; Also, it can analyze the rich archived video databases such as the TV broadcasted sports game videos or the videos recorded during dairy training. This paper presents our work on the automatic analysis of complex athlete action in diving video sequences.

However, the automatic analysis of sports action is not an easy task and most of the current work are either limited to purposefully recorded videos or need manual intervention. For example, Pascual J. *et al.* [1] developed a method for soccer player's position tracking aiming at their kinematical motion analysis through a graph representation with four static cameras. Though their method can collect statistics measurements of each player in the game, manual tracking is needed in some cases. J.R. Wang *et al.* [2] classified tennis games into 58 winning tactics patterns for archiving video clips and training purpose using the ball trajectory and bayesian network. To recover the trajectory and ball landing position they turned to a wide-view calibrated camera. The widely-studied human body model based tracking approach has also been suggested for sports biometric analysis. As Andrew *et al.* showed in [3], a 42 dimensional body model was used to track the golfer's postural information and then analyzed it with respect to a learned ideal motion. However, their system ran so slow that each frame would take approximately 25 minutes to process, and the initial parameters of body model should be set up manually before tracking. The most similar work to our diving action analysis was provided by Cassel *et al.* [4]. In their system, Cassel *et al.* analyzed the acrobatic gestures of several sports through modeling and characterizing acrobatic movements and image processing techniques. Though their system worked automatically, it was less challenging since that the athlete movements were captured with static camera and only simple acrobatic gestures were recognized by global measurements analysis.

In this paper, we propose a system for automatically analyzing complex diving action in challenging dynamic background based on our previous work on athlete body segmentation and action recognition [5]. The aim is to automatic obtain the main biometric measurements such as the knee joint angle and hip joint angle during take-off period and the entry angle during entering period and to provide visual analyzing tools for quantitative analysis or perceptible comparison of athlete performance. The knee and hip joint angles are obtained by 2D articulated human body model fitting which initial parameters are transferred

14th International Conference on Image Analysis and Processing (ICIAP 2007)
0-7695-2877-5/07 $25.00 © 2007 IEEE

from the recognized action templates automatically. The entry angle is got using shape analysis technique based on the accurate segmentation of athlete body. Two useful visual analyzing tools: motion panorama and overlay composition are also presented.

The rest of the paper is organized as follows. Our previous work on the athlete body segmentation and action recognition are first briefly reviewed in Section 2. Then the proposed biometric and visual analysis are presented in Section 3 and Section 4 respectively. Experimental results are given in Section 5. Finally we conclude the paper in Section 6.

2. Athlete Body Segmentation and Action Recognition

A robust object segmentation algorithm based on adaptive dynamic background construction and a hidden markov models (HMMs) based action recognition method have been proposed in our previous work [5]. For a given video clip containing the diving action, the background image for each frame of the clip is first built by adaptive selecting and aligning multiple neighboring keyframes to the current frame and then background subtraction technique is applied to segment the athlete body. With the segmented body shape sequence, continous HMMs are used to identify the action category. More details on the segmentation and recognition algorithms are referred to [5]. These algorithms enable the following work on the analysis of athlete action in this paper.

3. Automatic Biometric Analysis of Athlete Action

Biometric information is very useful for coaches to instruct athletes more scientifically. For diving action, the hip angle and knee angle in the takeoff period are of the most critical biometric parameters, since that the extent of flexion of hip and knee decide the height of the diving. Generally, a higher dive means longer time in the flight for the diver to hold the dive position longer and have more time to prepare for entry. A long flight time normally results in an aesthetically pleasing dive and, as a result, high scores. In the period of entry, the entry angle is another important factor to measure the performance of the dive since a vertical or near vertical entry will lead small splash than a slant entry, resulting a high score. The aim of biometric analysis in this paper is to provide biometric measurements for sports professionals to quantitatively analyze the action. Articulated human model fitting and silhouette analysis techniques are exploited to get the body angles in the takeoff and entry period respectively.

3.1 Human Body Model Fitting

Fitting an articulated human body model to image cues is the popular approach to obtaining the posture information such as body joint angles or joint locations [6]. In our case, a 11 degrees of freedom (DOFs) Scaled Prismatic Model [7] represented as $S = (x, y, \theta, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5, d_1, d_2, d_3)$ is adopted. In *S*, (x, y, θ) is the global position (the hip joint) and rotation parameters, $(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ are the left hip angle, left knee angle, right hip angle, right knee angle and neck angle respectively, and (d_1, d_2, d_3) are the length DOF for trunk, left leg and right leg.

The limitations of model-based human tracking are that it relies on manual initialization of the model parameters before the tracking to start, and often loses tracking due to accumulative error. Our approach circumvents these two disadvantages through action recognition. This is done by manually labeling the 2D joint locations of the training sample shapes for each action in the training stage. In the testing stage, when a test sequence is recognized, the corresponding hidden states are also decoded using viterbi algorithm and thus the initial model parameters are transferred from the sample shape of these states. The global position, to say, the hip joint is located as the mass center of the segmented body shape. When initialized, the final model parameters are searched with the well-known annealed particle filtering algorithm [8]. Since the initial parameters for each body shape are obtained from the recognized action templates independently, our approach has not the problem of losing tracking.

When matching the body model to the image, a multi-cue observation model which considers the shape and foreground area is adopted to make the matching robust and efficient. For shape, we use the distance transform [9] as similarity measurement between image edge features and model edge features. The image edge map is first filtered by removing the edges outside of the bounding rectangle of the predicted body and the edges inside the foreground area (segmented body area). Let *d* be the mean sum of each edge's cost, the shape matching score for model hypothesis x_i is:

$$p(z_{shape} \mid x_i) = \exp(-\lambda_1 * d) \tag{1}$$

For foreground, the model's overlapping rate *r* to the foreground area is first compute, then the foreground matching score is defined as:

$$p(z_{foreground} \mid x_i) = \exp(\lambda_2 * r) \tag{2}$$

Where λ_1 and λ_2 are two constants and are set with 10 and 1 respectively in our experiments.

The total matching score is:

$$p(z \mid x_i) = p(z_{shape} \mid x_i) * p(z_{foreground} \mid x_i) \quad (3)$$

The human body model and an example of fitting are shown in Fig. 1.

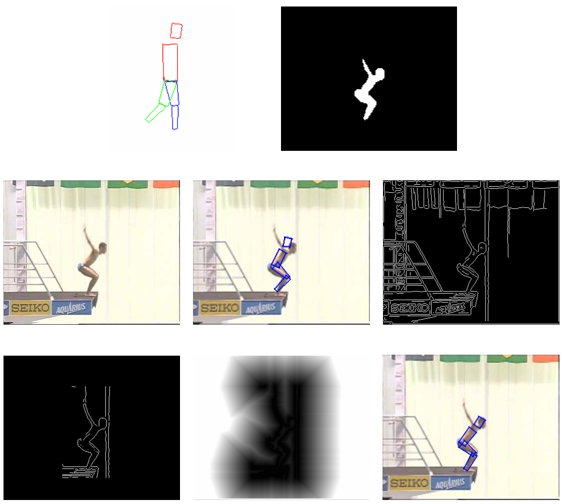


Fig. 1. Human body model and an example of model fitting result. From the first to the last image are the human body model, a test body shape, original frame, the initialized model, canny edge map, filtered edge map, distance transform map and the final fitted model respectively.

3.2 Silhouette Analysis

To analyze the entry angle, we need detect the entry point firstly. In the diving community, the entry point is defined as the first frame in which the hands have broken the water. If the gap between the hands and the water is very small and the next frame has the divers arms immersed almost up to the shoulder then the frame just prior to breaking the surface will be used as the entry. We first employ a learning-based method to detect the frame near entry point and then analyze the segmented shape to determine the true entry point.

As demonstrated in Fig. 2, the diver will try his best to extend his body before hitting the water to obtain an elegant entry; and when he is hitting the water, the broken water will be segmented out showing a dilated area at bottom of the segmented body shape. We collect these shape sub-sequences (3 frames are used in our experiments) in training stage and model the distribution of concatenated shape descriptor [5] with GMM. When detecting for a action clip, a three-frame sliding window is checked from the frame when the swimming pool appears in the image and the body shape is extended (with a large Hu1 value above 0.5 [5]). The frame with the maximum GMM output probability is selected as the approximate entry point (this will likely be the image (d) in Fig. 2). We then analyze the segmented shapes of the several frames near the detected entry point which have dilated bottom ends (see images in Fig. 2 (e)-(f)) to find the true entry.

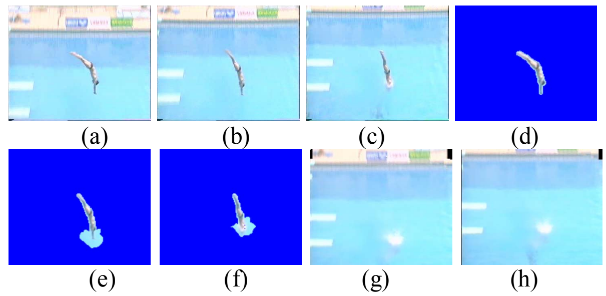
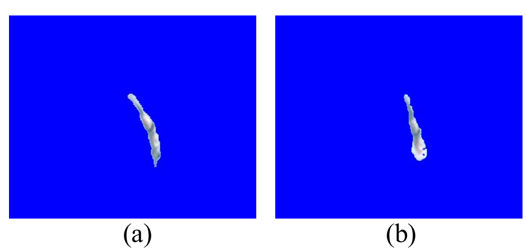


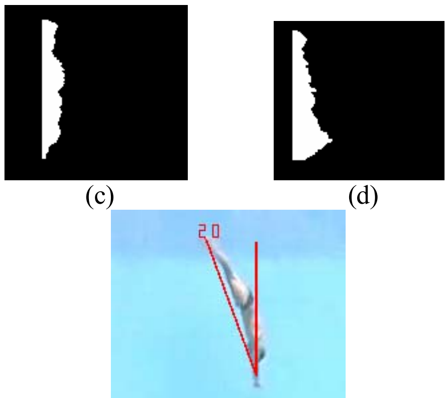
Fig. 2. Three frames near the entry point and the respective segmented foregrounds: (a)-(c) are the frames, (d)-(f) are the segmented images and (g)-(h) are the constructed background images of (d)-(e).

The pool color filtering is used to remove the segmented water in the body shape which is caused by the spray in the constructed background. Fig. 3 (a)-(b) show the filtered images of Fig. 2 (e)-(f). Then the horizontal projection of the filtered silhouette is computed to determine the true entry. As shown in Fig. 3 (c)-(d), the projection value of the bottom 1/3 part of the entry frame is smaller than that of the middle 1/3 part, while it is converse for the following frame. So a true entry point *t* should satisfy:

$$\begin{cases} P_B(t) < P_M(t) \times 1.2 \\ P_B(t+1) > P_M(t+1) \times 1.2 \end{cases} \quad (4)$$

Where $P_B(t)$ is the mean projection value of the bottom 1/3 part for frame *t* and $P_M(t)$ is the mean projection value of the middle 1/3 part for frame *t* (if no frame satisfies (4), then the firstly detected entry point is adopted). In this example, the image in Fig. 3 (a) is detected as the entry point. After the entry is detected, the most far points to the mass center above and below it in the foreground image are selected and the angle from the linking line of the two points to the vertical line is defined as the entry angle (Fig. 3 (e)).





(c) (d)

(e)

Fig. 3. The filtered foregrounds(a-b), their horizontal projections(c-d) and the detected entry angle.

4. Visual Analysis of Athlete Action

With the enabling techniques, to say, the accurate athlete body segmentation and action recognition, we present two visual sports training tools in this paper which allow coaches and athletes to review and compare performances in an easy way to enhance coaching strategies.

4.1 Motion Panorama

Motion panorama provides an efficient and compact representation of the underlying video by constructing a single image from part of frames of a video. It can show the global background as well as the foreground athlete bodies and is more straightforward to reveal the moving trajectory and action details than by a video. Panorama has been proved to be a powerful tool for reviewing and evaluating the performance of athlete action [10].

After estimating global motion parameters, a background panorama is built first. All video frames are aligned to the coordinate system of the first frame, which is also selected as the world coordinate system of the panorama. Then for those overlapping regions, temporal median filtering technique is used to construct the background by selecting the median RGB value.

Using the global motion and time intervals as criterions, some key-frames are chosen automatically, of course a manual selecting function is also provided. The segmented foregrounds of such key-frames are mapped to a world coordinate system same to the background panorama to form a foreground panorama. For visual effect, the overlapped foreground regions are fused by alpha blending.

Finally, the resulting motion panorama is created by covering the background panorama with the effective regions of foreground panorama. An example of panoramas for a platform diving clip is shown in Fig. 4.

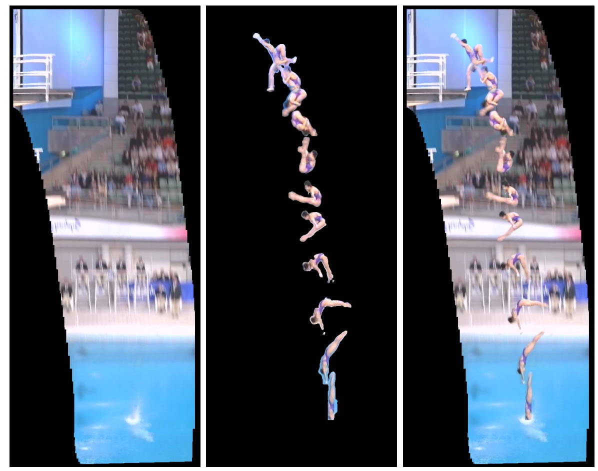


Fig. 4. The foreground panorama, background panorama and motion panorama.

4.2 Overlay Composition

Overlay composition is also a video blending technique which creates a compositive video by overlaying two temporally aligned videos at the same scene [11]. It provides a visually appealing tool for coaches and athletes to see minute changes between same action performed by different athletes or by the same athlete at different time. This is important for sports training since that very small differences in technique can turn into very large differences in performance in competitive sports games.

Given a synchronization point of two clips and the segmented foregrounds, we compose an overlay clip by superimposing one clip's foregrounds over the other clip using alpha blending technique. Unlike the traditional overlay method, our method does not have the limitation of requiring same scene for the two clips, since that we have segmented the foreground accurately; Also the synchronization point can be automatically determined by temporally aligning the shape sequence with techniques like dynamic time warping. In our case, we take the entry point as synchronization point to compose the overlay clips for the same action. Fig. 5 gives two example frames of an overlay clip.

Fig. 5. Overlay frames of same action performed by two athletes from two clips

5. Experimental Results

This section reports our experimental results on the analysis of diving videos, digitized in MPEG 1 format. The videos include platform videos and springboard videos and they come from 27th Olympic Games (Sydney, 2000) and 9th World Tournament (2001, Fukuoka). The ground truth is labeled manually.

5.1 Biometric Analysis

The human body fitting is used to obtain athlete's biometric information, to say, hip angle and knee angle in the takeoff period. Four kinds of action: 205B, 305C, 307C and 5251B, are selected to test the algorithm. The main body joints (head, shoulder, hip, knee and foot) of the training shapes of these actions in takeoff period are labeled manually with a friendly GUI. And then the body model parameters are computed. When the HMM for an action is trained, we compute the mean body model parameters for each Gaussians of the hidden states. In testing, the HMM states are decoded using viterbi algorithm and thus the initial body model parameters for each frame are obtained. Finally, the annealed particle filtering with 500 particles and 3 layers is used to refine the model parameters. The errors for hip angle and knee angle between the manually labeled and automatic tracked are listed in Table 1. It can be seen that the automatic method has achieved comparing precision with the ground truth and the measurements are precise enough for quantitative analysis of athlete's performance. Some more model fitting examples for action 307C are illuminated in Fig. 6.

Table 1. The errors for hip angle and knee angle.

	Mean error	Max. error	Min. error
Hip angle (degree)	1.32	5.35	0.06
Knee angle(degree)	0.88	5.71	0.04

The silhouette analysis technique is used to detect the entry point and to analyze entry angle. The shape distribution for the entry sub-sequences are trained and modeled using a GMM with 8 components, representing the 4 entry modes. In our experiments, the performance is evaluated by the diving professionals. For most cases we have achieved satisfying results for entry point detection and entry angle measuring. Fig. 7 gives some more results.

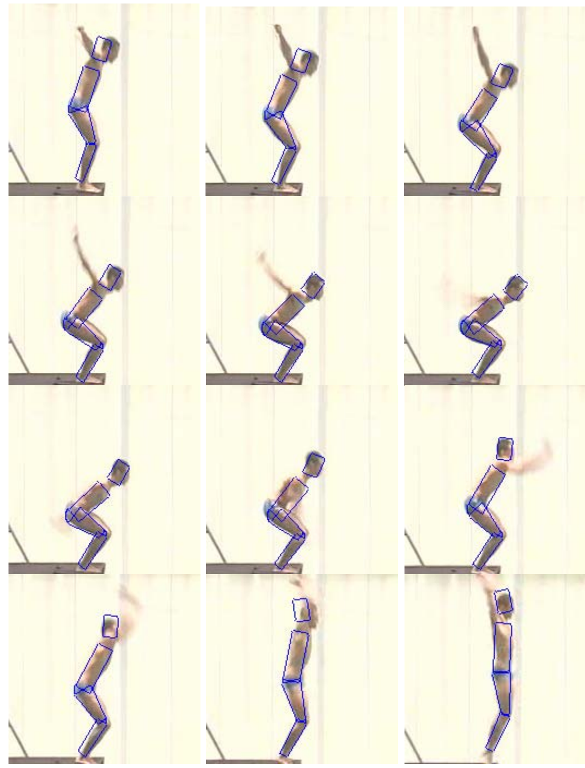


Fig. 6. Human model fitting for action 307C.

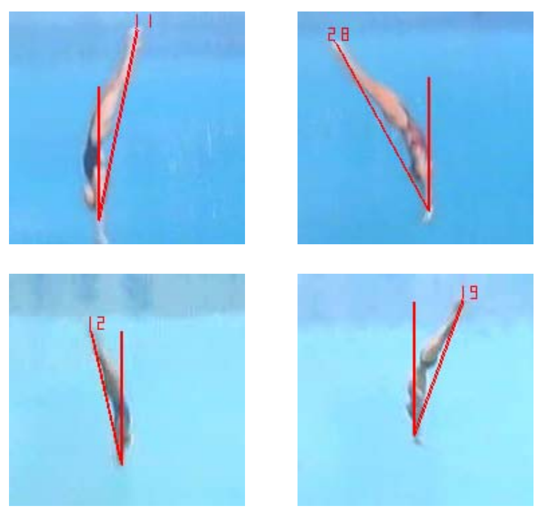


Fig. 7. Entry angle measuring

5.2 Visual Analysis

The presented visual analyzing tools: motion panorama and overlay composition are very suitable for the individual sports action analysis. Here we illuminate more results for motion panorama and overlay composition as in Fig. 8 and Fig. 9 respectively.

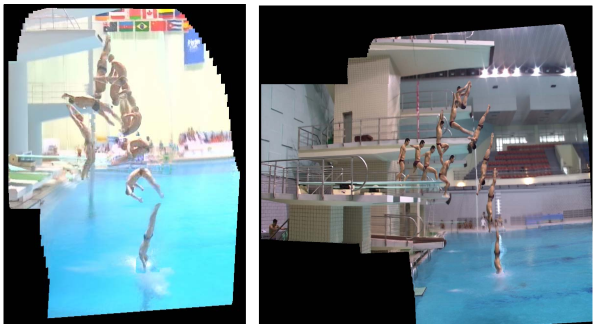


Fig. 8. Motion panorama for two springboard diving clips

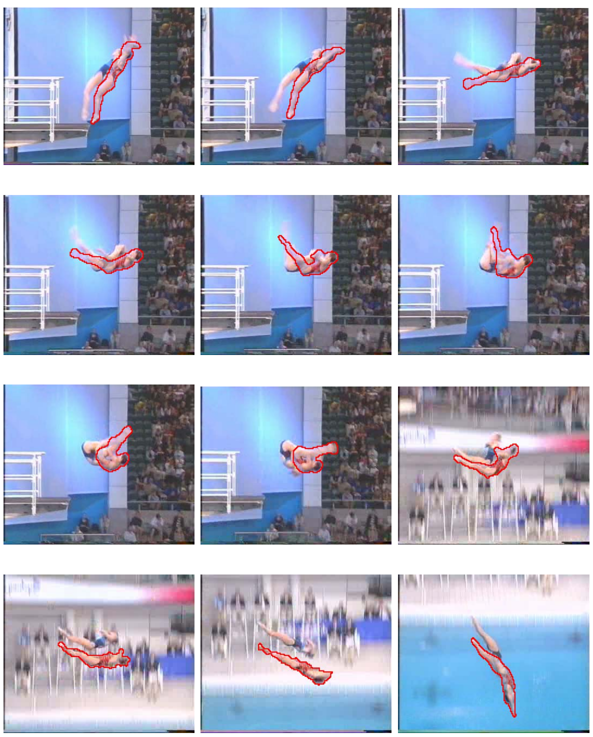


Fig. 9. Overlay composition for action 205B of two divers

6. Conclusion and Future Work

We have presented our work on the automatic analysis of athlete motion in action-critical sports video, to say, diving video for training purpose. Based on the accurate athlete body segmentation and action recognition, the automatic biometric analysis is achieved and two useful visual analyzing tools are provided. Though they are tested on diving video analysis in current work, it is easy to adapt them to analyze other kinds of sports videos, such as gymnastics, trampoline, and so on. In the future, we will extend our method to obtain more biometric measurements for training instruction.

Acknowledgements

This work is supported by Beijing Science and Technology Planning Program of China (D0106008040291)

References

[1] P. Figueroa, N. Leite and R.M.L. Barros, Tracking soccer players aiming their kinematical motion analysis, Computer Vision and Image Understanding, 2006, 101(2), pp. 122-135.

[2] Jenny R. Wang, Nandan Parameswaran, Detecting Tactics Patterns for Archiving Tennis Video Clips, in: Proc. ISMSE'04, Miami, FL, USA, 2004, pp. 186-192.

[3] Smith, Andrew W. B., Lovell, Brian C, Visual Tracking for Sports Applications, in: Proc. WDIC'05, Brisbane, 2005, pp. 79-84.3

[4] Ryan Cassel, Christophe Collet, Rachid Gherbi, Real-Time Acrobatic Gesture Analysis, Gesture Workshop'05, Berder Island, France, 2005, pp.88-99.

[5] Haojie Li, Si Wu, Shouxun Lin, Yongdong Zhang, Automatic Detection and Recognition of Athlete Actions in Diving Video, in Proc. MMM2007, Singapore, 2007, LNCS 4352, pp. 73-83

[6] TB Moeslund , Adrian Hilton , Volker Krüger, A survey of advances in vision-based human motion capture and analysis, Computer Vision and Image Understanding, 2006, 104(2), pp.90-126

[7] DD Morris, J. Rehg, Singularity analysis for articulated object tracking, in: Proc CVPR'98, Santa Barbara, CA, 1998, pp. 286-296.

[8] J. Deutscher, A. Blake, et al, Articulated body motion capture by annealed particle filtering, in: Proc CVPR'00, Hilton Head Island, 2000, pp. 126-133.

[9] D. M. Gavrila, V. Philomin, Real-time Object Detection for Smart Vehicles, in: Proc. ICCV'99, Kerkyra, 1999, pp. 87-93.

[10] Chiou-Ting Hsu, Yu-Chun Tsan, Mosaics of video sequences with moving objects, Signal Processing: Image Communication, 2004, 19, pp. 81-98.

[11] http://www.dartfish.com