Motion Analysis for the Standing Long Jump

Hui-Huang Hsu[1], Sheng-Wen Hsieh[1], Wu-Chou Chen[2], Chun-Jung Chen[3], and Che-Yu Yang[1]
[1]*Department of Computer Science and Information Engineering, Tamkang University*
[2]*Graduate Institute of Sports Science, National College of Physical Education and Sports*
[3]*Department of Computer Science, Chinese Culture University*
Taipei, Taiwan, R.O.C.
E-mail: h_hsu@mail.tku.edu.tw

Abstract

The standing long jump is a standard test for primary school students. It can be used to evaluate the development of basic sports skills of a child. This paper presents a system that can automatically detect the motion during a standing long jump from a video sequence. The silhouette of the jumper in the film is segmented from the background first for all frames. A stick model is applied to the silhouette found in the first frame. Then a GA-based search algorithm is used to find the stick models for the rest of the frames. The stick model points out the important joints of a person and can be used to represent the pose of the jumper in each frame. From the pose change in consecutive frames, we will be able to analyze the movement of the jumper.

Keywords: *motion analysis, human detection and tracking, shadow removal, pose estimation, genetic algorithms*

1. Introduction

In physical education, there are several standard tests for understanding the development of sports skills of children. For example, the standing long jump and ball throwing. The standing long jump can be observed from the side. Two-dimensional (2D) information is sufficient to decide the performance of the jumper. On the other hand, ball throwing involves 3D information for the evaluation. That would be a more difficult task. In this work, we aim at developing a system to evaluate the standing long jump from video sequences. There are three parts for such a system: (1) Human detection, (2) Pose estimation, and (3) Scoring. The system will be able to detect improper movements and give advices to the jumper.

Human detection has been an important research issue in computer vision because of its wide range of applications. Two things need to be done here. First, moving objects in a video sequence are segmented from the background. Secondly, shadows are removed. In [1], hybrid sensing of depth and gray information by a single camera is used to achieve 3D detection of human. It also presents the application on human tracking and event detection. However, the effect of shadows is not discussed. In [2], a change detector and a shadow detector are combined for object-based video processing. It also incorporates color information and spatio-temporal verification for human object segmentation. In [3] and [4], HSV color information is used to detect and suppress shadows. Moving objects can be segmented without the interference of moving shadows.

For pose estimation, a genetic algorithm (GA) is used to estimate the pose of a human object from its silhouette in [5]. With the GA-based search, a proper stick model with a high accuracy can be found in 200 generations. However, no temporal information is utilized. In this work, a modified version is developed for video sequences. With pose information, automatic scoring of the jump becomes possible. In physical education, standards for the standing long jump have been proposed. They can be used to analyze or score the movement of the standing long jump. Scoring rules associated with the stick model information are derived from these standards.

In this paper, the algorithms of the first two parts of the system are presented. In the following, the segmentation of human objects is described first in Section 2. Then the pose estimation method is discussed in Section 3. Scoring rules to identify improper movements are formulated in Section 4.

Proceedings of the 26th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'06)

The preliminary results are shown in Section 5. In Section 6, conclusions are drawn and the future work is indicated.

2. Segmentation of Human Objects

As mentioned earlier, human objects are segmented from a video sequence by first detecting moving objects and then removing shadows. The whole algorithm can be divided into five steps:

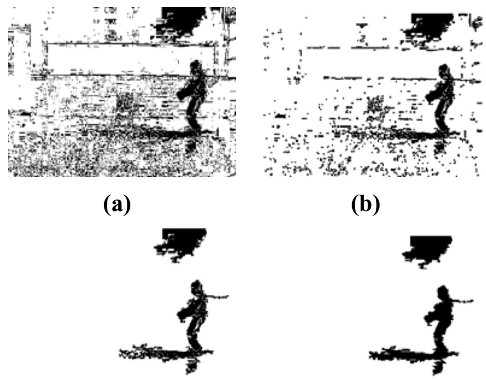
(1) Generate the background image for a video sequence;
(2) Subtract the background image from each frame;
(3) Remove noises and small spots caused by the light change from each frame;
(4) Fill up small holes in the objects;
(5) Remove shadows.

In the first step, the background can be estimated by change detection. The pixels with a very small change in two consecutive frames are saved as part of the background. This process goes from the first two frames to the final two frames in the video sequence. In Figure 1, the first frame of a video sequence and the generated background are shown.

(a) (b)

Figure 1. (a) First frame of a standing long jump video sequence, and (b) the estimated background from the video sequence

In Step 2, the background is subtracted from each frame to obtain the foreground of each frame as in Figure 2 (a). A lot of noise due to light changes can be seen. In Step 3, noises are deleted by checking the eight neighbors of a pixel in a frame. If the number of neighbors that are not 0 is greater than the threshold, the pixel is kept. Otherwise, the pixel value is set to 0. Figure 2 (b) can be obtained after this. And, since we are looking for human objects, smaller spots can be removed from the scene. Figure 2 (c) is the result. In Step 4, small holes are filled up in the remaining objects. If a pixel in the object is 0 and the four neighbors of the pixel are all 1, the value of the pixel is set to 1. Otherwise, it remains 0. Figure 2 (d) shows the result.



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

Figure 2. Results of foreground extraction after (a) subtraction of the background (b) removal of noises (c) removal of small spots (d) fill-up of small holes

In Step 5, the method for shadow detection presented in [3] and [4] is used. The pixels are analyzed in the Hue-Saturation-Value (HSV) space. If a shadow is cast on the background, both the hue component and the saturation component change. However, the difference in saturation is an absolute value and the difference in hue is an angular value. A shadow mask (SM_k) for each pixel (p) at frame k, to be applied to the extracted objects, is defined as follows with three conditions.

$$SM_k(p) = \begin{cases} 1 & \text{if } \alpha \le \frac{F_k(p).V}{B_k(p).V} \le \beta \ \wedge \ |F_k(p).S - B_k(p).S| \le \tau_S \ \wedge \ DH \le \tau_H \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $F_k(p)$ is the vector representation of the segmented object in the HSV space, $B_k(p)$ is the vector representation of the background in the HSV space, and DH is computed by

$$DH_k(p) = \min(|F_k(p).H - B_k(p).H|\ ,\ 360 - |F_k(p).H - B_k(p).H|)\ . \quad (2)$$

Also, $F_k(p).H$, $F_k(p).S$ and $F_k(p).V$ are the hue component, the saturation component, and the value component of segmented point $F_k(p)$, respectively. $B_k(p).H$, $B_k(p).S$ and $B_k(p).V$ are the hue component, the saturation component, and the value component of background point $B_k(p)$, respectively. The parameters - α, β, τ_S, and τ_H are determined via experiments. Figure 3 is obtained

Proceedings of the 26th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'06)
0-7695-2541-5 /06 $20.00 © 2006 IEEE

from Figure 2 (d) with shadows removed. Comparing Figure 3 (b) with Figure 1 (a), we can see that the result for human segmentation is quite successful.

(a) (b)

Figure 3. Results of shadow removal (a) in black (b) in original colors

3. Pose Estimation

To estimate human pose from a silhouette is not an easy task. In [5], a genetic algorithm (GA) was developed to handle the pose estimation for a single silhouette. It has to assume a stick model with a known size compatible to the silhouette first. Here we have a sequence of silhouettes. In order to determine an appropriate size for the stick model, a trained person is asked to draw the stick figure for the human object in the first frame. Then a GA-based algorithm modified from [5] is used to find the pose of the human object in each of the rest frames. The stick model for the standing long jump is shown in Figure 4. Since the video is always taken from the side, two arms and two legs are merged into one, respectively. In the stick model, S_0 is the trunk, S_1 is the neck, S_2 is the upper arm, S_3 is the thigh, S_4 is the head, S_5 is the forearm, S_6 is the shank, and S_7 is the foot.

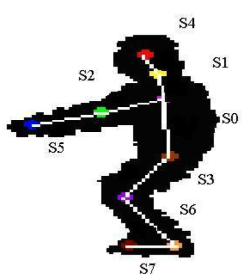


Figure 4. Stick model for standing long jump

The stick model can be represented by the coordinates of the center (x_0, y_0) of Stick S_0 and the angles (ρ_l) of Stick S_l ($l = 0, 1, \ldots, 7$) with the vertical line (y axis):
$(x_0, y_0, \rho_0, \rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6, \rho_7)$.
The angle (ρ_l) associated with a stick (S_l) is illustrated in Figure 5. The lower end of S_0 and the one end of S_l ($l = 1, \ldots, 7$) nearer to the trunk are placed at the origin. And the angle from the y axis to Stick S_l is Angle ρ_l associated with it.

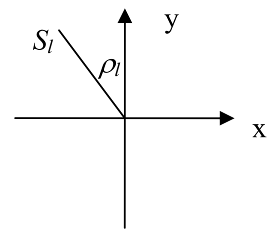


Figure 5. Angle of a stick in the stick model

To use a GA to search for a proper stick model for each silhouette, four things are needed:
(1) Representation of the stick model as a chromosome,
(2) A fitness function,
(3) An initial population of chromosomes, and
(4) An evolution strategy.

First, the stick model data can be represented as $(x_0, y_0, \rho_0, \rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6, \rho_7)$, where x_0 and y_0 should locate near the center of mass of the silhouette, and ρ_l ranges from 0° to 360° as long as the corresponding stick is in the silhouette. The representation is called a chromosome in the GA.

Next, a fitness function is needed to evaluate the fitness of a certain stick model in a silhouette. In the GA, only the fittest chromosomes can survive. A fitness function is as follows.

$$F_S = \frac{\sum_{i,j} \min_{l=0,\ldots,7}\{d_l(x_i, y_j)/t_l\}}{N} \quad (3)$$

where (x_i, y_j) is any point in the silhouette, t_l is the average thickness of the area surrounding Stick S_l, and N is the total number of points in the silhouette. The thickness of all sticks' area can be estimated from the stick model drawn by human in the first frame. In this fitness function, the smaller the F_S is, the better the stick model fits the silhouette.

To estimate the stick model for the second frame, the center of S_0 (x_0, y_0) is put at the geometric center of the silhouette (x_c, y_c). Points from the rectangle $\{(x_c-\Delta x, y_c-\Delta y), (x_c+\Delta x, y_c+\Delta y)\}$ can be randomly selected to be in the initial population. For each angle ρ_l, the corresponding angel in the stick model of the preceding frame is ρ_i^{-1}. The initial angles can be randomly chosen from the range of $\rho_l^{-1} \pm \Delta\rho_l$, where $\Delta\rho_l$ is different for

Proceedings of the 26th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'06)
0-7695-2541-5 /06 $20.00 © 2006 IEEE

different sticks. $\Delta\rho_l$ can be determined by the nature of connected joints of the corresponding stick. Any randomly-generated chromosome not in the boundary of the silhouette should be removed from the initial population.

For the evolution strategy, the elitism is used. Meaning, in each generation, only the fittest chromosomes can be left and they have a higher probability to be picked for generating the next generation. Crossover and mutation are applied to two selected chromosomes to generate new chromosomes. Multiple crossover is used with genes in the chromosome grouped as follows: (x_0, y_0), (ρ_0), (ρ_1, ρ_4), (ρ_2, ρ_5), (ρ_3, ρ_6, ρ_7). The angles of the sticks of neck-head and four limbs are put together, respectively. We can set the crossover rate to 0.2. After a crossover, mutation can be applied to each group with a probability 0.01. In this phase, the generated chromosomes not in the silhouette are also removed from the population.

4. Scoring Rules

From the discussion with physical education experts, standards to evaluate the standing long jump are formulated. Table 1 shows some of the sample standards for initiating the project.

The standards in Table 1 can be translated into scoring rules of the stick model data. Table 2 lists the evaluation rules corresponding to the standards in Table 1. The rules can be used to implement the scoring component of the system. The angles are part of the stick model formulated in Section 3. However, it is necessary to examine the angles for a few consecutive frames in order to decide if the rules are satisfied. For example, to check R1, the angle difference between ρ_6 and ρ_3 should be examined from the first frame to the 10th frame and the maximum of all the angle differences is then used. To check R6, the angle of ρ_0 should be examined from the 11th frame to the 20th frame and the maximum is then used. These rules are only preliminary. Further discussions and enhancement of the rules is necessary.

Table 1. Standing long jump evaluation standards

Initiation Stage	
E1:	Knees bended
E2:	Neck bended forward
E3:	Arms swung back
E4:	Arms bended
On the Air/Landing	
E5:	Knees bended
E6:	Trunk bended forward
E7:	Arms swung forward after landing

Table 2. Scoring rules for standing long jump

Initiation Stage	
R1:	$\rho_6 - \rho_3 > 60°$
R2:	$\rho_1 > 30°$
R3:	$\rho_2 > 270°$
R4:	$\rho_2 - \rho_5 > 45°$
On the Air/Landing	
R5:	$\rho_6 - \rho_3 > 60°$
R6:	$\rho_0 > 45°$
R7:	$\rho_2 < 160°$

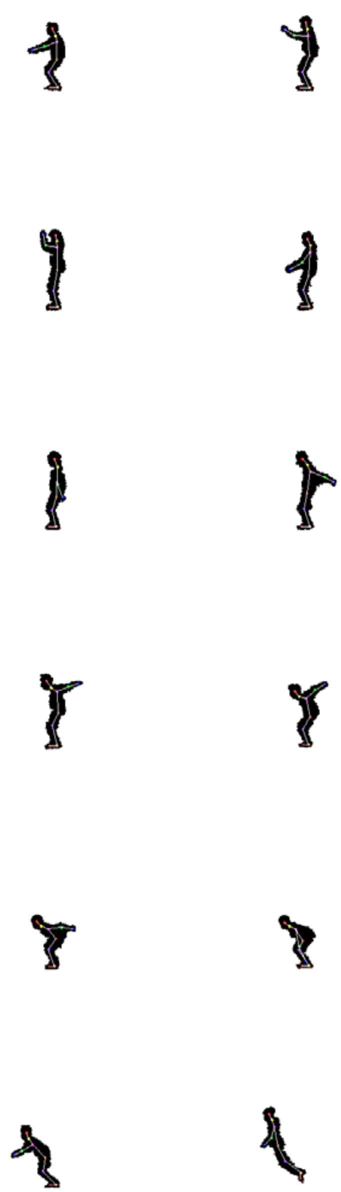
5. Preliminary Results

In this section, the preliminary results for estimating the silhouettes of the human object in a video sequence and their corresponding stick models are presented. There are totally 20 frames or so for a standing long jump video sequence. Samples of computer-generated silhouettes and the manually-drawn stick models in a standing long jump video sequence are shown in Figure 6. With the estimated stick models in consecutive frames, we will be able to track the movement of the jumper and to use the scoring rules in Table 2 to evaluate the movement.

Figure 7 shows the estimated stick models for the second and third frames of the example in Figure 6. The two estimated model were generated automatically by the developed system. The results

Proceedings of the 26th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'06)
0-7695-2541-5 /06 $20.00 © 2006 IEEE

were quite good. The initial population for estimating the second frame was derived from the first frame. And the shown best estimated model was generated at the second generation. Similarly, the initial population for estimating the third frame was derived from the second frame. And the shown best estimated model was also generated at the second generation.



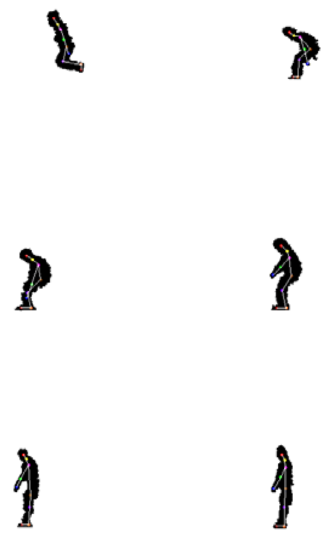


Figure 6. Silhouettes and manually-drawn stick models of consecutive frames

Figure 7. Silhouettes and computer-generated stick models of the second and third frames in Figure 6

6. Conclusions and Future Work

We have developed a system for motion analysis of the standing long jump though the scoring part is yet to be implemented and tested. The preliminary results are quite promising for final development of an automatic scoring system. The scoring system will soon be developed and the

Proceedings of the 26th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'06)
0-7695-2541-5 /06 $20.00 © 2006 IEEE

results will be compared with human evaluation. In the future, we would also like to build a web-based system on the Internet. The user will be able to upload a video sequence of a standing long jump or capture the standing long jump with a CCD camera online. With a proper setting of the video capturing, the system will be able to respond with advices to the user. Also, the more complicated task of recognizing ball throwing motion will be challenged.

References

[1] Fengliang Xu and Kikuo Fujimura, "Human Detection Using Depth and Gray Images," *Proc. IEEE Conference on Advanced Video and Signal Based Surveillance*, 115 – 121, July 21-22, 2003.

[2] A. Cavallaro, E. Salvador, and T. Ebrahimi, "Shadow-aware object-based video processing," *IEE Proc. - Vision, Image and Signal Processing*, 152 (4), 398 – 406, Aug. 5, 2005.

[3] Rita Cucchiara, Costantino Grana, Massimo Piccardi, and Andrea Prati, "Detecting Moving Objects, Ghosts, and Shadows in Video Streams," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25 (10), 1337 – 1342, Oct. 2003.

[4] Rita Cucchiara, Costantino Grana, Massimo Piccardi, Andrea Prati, and Stefano Sirotti, "Improving Shadow Suppression in Moving Object Detection with HSV Color Information," *Proc. 2001 IEEE Intelligent Transportation Systems Conf.*, 334 – 339, Aug. 25-29, 2001.

[5] Kenji Shoji, Atsushi Mito, and Fubito Toyama, "Pose Estimation of a 2D Articulated Object from its Silhouette Using a GA," Proc. 15th Int'l Conf. on Pattern Recognition, 3, 713 – 717, Sept. 3-7, 2000.

Proceedings of the 26th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'06)
0-7695-2541-5 /06 $20.00 © 2006 IEEE