Tracking of Real Time Acrobatic Movements by Image Processing.

Ryan Cassel and Christophe Collet

Dept. of Communication Homme-Machine groupe Geste & image
LIMSI-CNRS, University of Paris XI, BP 133, 91403 Orsay cedex, France
{ryan.cassel, christophe.collet}@limsi.fr

Abstract. This paper presents the design and evaluation approach of a video camera-based system for tracking and evaluating acrobatic movements. The context of this study is human gesture recognition in trampoline competition performance. Our work builds a model of trampoline movement and conducts efficient real time tracking of the trampolinist in order to recognize and evaluate a competition routine. We describe the prospective architecture of our system, which satisfies constraints of real time image processing. The global project includes three consecutive phases: first, body extraction and tracking in the image sequence, second, body part localization, and last, gesture recognition and quantification according to our model of trampoline movement characterization. This paper describes the first phase, which combines image processing techniques to perform fast and efficient extraction and tracking of the trampolinist's body. We then set out an evaluation protocol and the results for this phase.

1 Introduction

Studies concerning the analysis and evaluation of movements based on sports are of growing interest for researchers and trainers. Researchers want to understand how the body performs sporting movements. Trainers want to teach these movements and to improve their realization. Most studies are based on video analysis performed by experts or direct tracking of the body using specific captors. The use of magnetic or infra-red captors enables precise measurements with very high accuracy and frequency, but movements are limited and less natural [1]. In this context, the use of a video based automatic gesture tracking and recognition system becomes very helpful to perform this type of analysis [2]. These system use image processing techniques to effectuate measurement in the video sequence. But their fields of application are limited movements or simple gestures (walking, dancing, designation... [3] [4]). Movement characterization represents a major problem. Many studies focus on segmentation of body parts, human model and complex system to translate basic movements [5] [6] [7]. These algorithms could be more efficient if the observed movement were more clearly characterized. Our purpose is to build a single camera-based system, using a general model with few points to find and track a body in real time. In the proposed method, we characterize and track motion in order to recognize complex acrobatic gestures. We choose to apply this study to the recognition and evaluation of trampolining motion. Trampoline offers several complex body gestures, specifically characterized movements, and a taxonomy of these movements. In trampoline training as well as in competition, gestures need to be analyzed in real time. Real time computer vision for use in sports has to use computationally inexpensive algorithms, be generally sufficient for work in various environments without significant parameter tweaking, and comprise automatic start and initialization with minimum necessary knowledge about the observed scene [8]. The architecture of the system can be segmented as follows. First, we detect and track the human body to create a bounding box. In a second step, we identify each body part using a general model (head, body axis and feet see figure 2). Finally, the body configuration is analyzed dynamically to recognize body gestures.

This paper deals with the characterization of human gestures on a trampoline, then exposes fast tracking algorithms that prepare relevant data for real time recognition. Last, we present an experimental evaluation of this tracking algorithm.

2 Model of Trampoline Movement Characterization

Trampoline is a sport where the movement is defined to be recognized efficiently. An element is defined by salto and twist quantity. Using the FIG (Fédération Internationale de Gymnastique) numeric system, salto is divided by quarters and twist by halfs. In the *Code of points* [9] used for competition, an element can be represented by an alpha numeric value (for example: "8 0 0 o" is a double back tuck). The first digit describes the number of somersaults, in quarters ($^1/_4$). Subsequent digits represent the distribution and quantity of half twists in each somersault. The shape of the element (called position) is described at the end using an '*o*' or leaving a blank for tucked; '<' for pike and '/' for straight also called layout. In all positions, the feet and legs should be kept together and the feet and toes pointed. Depending on the requirements of the movement, the body should be either tucked, picked or straight (figure 1). This *Code of points*

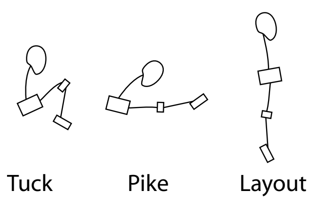

Fig. 1. Human body's positions in trampoline.

allows the judge to differentiate each element. Nevertheless, the judge looks at the trampolinist's routine and adds important information to the notation such as direction and the starting posture (from feet, back. . .) of the elements. For example, the numeric notation ”4 0 o” fails to indicate whether the figure is a backward somersault or a forward somersault. The information we do have is that the figure is a 4 quarters somersault, 0 half twist, in a tuck position. Therefore, the official *Code of points* is insufficient and ambiguous. It needs improvement if it is to be used as a model in an automatic recognition system. We add information about the direction of somersaults (forward or backward) at the beginning of the notation in order to be exhaustive for recognition purposes. A backward tuck somersault becomes ”b 4 0 o”, whereas a front tuck somersault becomes ”f 4 0 o”. We don’t add information about the starting posture because it can be deduced from the previous figure. All ambiguities are now lifted. In terms of data quantity, we conclude that we don’t need the entire body information.

We evaluate the body position by measuring angles between different body segments. The most important angles are: lower leg and thigh, thigh and torso. Thus, we need three body segments. To describe segments we use joints [10] for the head, the base, the knees, and the feet. For this study, we use a simplified model (figure 2). The minimum requirements for a particular body shape are defined as follows [9]:

- **Tuck position** : The angle between the upper body and thighs must be equal to or less than 135° and the angle between the thighs and the lower legs must be equal to or less than 135°.
- **Pike position** : The angle between the upper body and thighs must be equal to or less than 135° and the angle between the thighs and the lower legs must be greater than 135°.
- **Straight position** : The angle between the upper body and thighs must be greater than 135°.

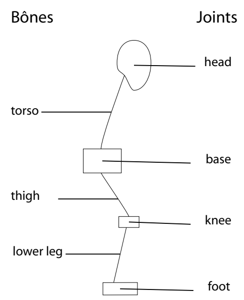

Fig. 2. Model composed of 3 segments and 4 joints.

The important period in the sequence is the ten competition elements. The sequence can be divided into two important periods and one final important signal.

In the initial period, the first jumps of the trampolinist are used to take dash. The second period (and the most important) is the ten competition elements. The final important signal is when the trampolinist stops his movement (he stops on the trampoline).

In order to comply with the international competition rules, the subject has to take amplitude that effectively leads to discovering the entire background. This information is crucial because an adaptive background will be generated.

3 Software Architecture of the Recognition System

To compute a recognition we need the subject images and the first step allows us to extract and track this data. In this article, we develop the tracking part; the subsequent steps (localization of body parts and recognition of body gesture) are still under investigation . The techniques to extract the subject image are successive image subtraction, background generation and motion filter for tracking.

Using the characterization described above, we need to detect, the direction, position, number of somersaults and the number of half twists depending on time. Immediately after capturing this data, the numeric notation can be sent to the judges. A coach can also analyze the starting twist in the somersault for comparison purposes.

4 Fast and Efficient Tracking Algorithm

In our method, image subtraction combined with a threshold are used to detect the most important motions. With each frame, the system generates an updated background image. Because of potential background changes (movements in the audience, light variations, shadow casting. . .), adaptive background is better than a method pulling a background image. However this method has not been evaluated yet.

The tracking algorithm is divided into three modules: a module to extract important motion parts (only the trampolinist), a module to automatically generate an adaptive background image, and a module to compute motion quantity. These three modules work first in an initialization phase. Once the background is generated, these modules continue in a processing phase to track the trampolinist and to extract a bounding box that contains the subject from the image sequence.

4.1 Subject Detection

In this context, the subject is always in the foreground and he is represented by a collection of neighboring pixels. The motion is extracted using a classical image-subtraction method. A dynamic threshold eliminates noise and highlights the main motion [11].

4.2 Filtering Motion

The notion of motion quantity is presented in [11, 12]. It is a value which characterizes a quantity of movement in the image, in order to determine the location of the bounding box according to its previous coordinates. For insignificant movements, the value of the motion quantity ($Q(t)$) is near 0; consequently, the system takes coordinates from the previous location of the bounding box. When significant movements are measured, $Q(t)$ is near 1; the location of the bounding box has to move to the new computed location. The advantage of this method is the stability of the bounding box and its computationally inexpensive algorithm.

4.3 Background Generation

Adaptive background generation [11, 13] allows the system to perform more efficient motion detection. Its principle is to generate a continuous background image that takes changes into account (light, shadows, audience...). A standard subtraction algorithm (operating between the generated background and the current image) detects only the trampolinist. To constitute a continuous background, the system takes an average value of each pixel which is not in the bounding box. An initialization phase allows the system to set up and generate this background.

4.4 Principle of the Tracking System

Background generation is triggered during the initialization phase, and this phase continues until the background image is completed. The tracking system is effective to track the trampolinist and return a valid bounding box. A subtraction image is computed from current and the adaptative background images. The result is added with the classic successive subtraction image that allows to highlight motion and the subject's body parts.

5 Experimental Results

To verify our proposed method, we tested the system with image sequences for three different trampolinists. The motion of the human body was taken with a Mini DV video camera, which can take 25 frames per second at 720 x 576 pixels. There is an average of 1700 images per sequence. The first sequence was used to set up the system, the two other was used for testing. The evaluation was done on recorded image sequences.

5.1 Protocol

We developed three software programs to evaluate our method [11]:

The first software is the main program described in this article which computes and returns the bounding box.

The second is an indexation software program to manually add information for each image of the sequence (to determine the coordinates of each body part). We manually labeled the three image sequences. The labeling consists in indexing all visible body parts to identify them and in recording each body part coordinates. We indexed the joints: the head, the base, the two hands, the two knees and the feet.

The last program takes the output of the two previous software programs and performs the evaluation. It computes statistics about all body parts in/out of the bounding box, frame by frame and measures the bounding box effectiveness for each video segment.

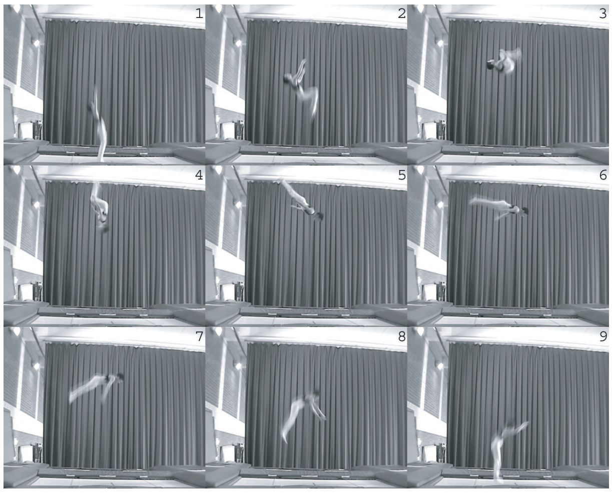

Fig. 3. Chronogram of a backward somersault (b 4 0 o).

To perform the evaluation, we started the main program to compute the bounding box coordinates. Then, we computed the belonging of body parts to the bounding box with the evaluation program.

5.2 Result

Table 1 presents the result of a normal course of the system, for the entire sequence and for the ten elements of the competition routine. The neutrality of hand movement during the first jumps contrary to the 10 element period explains the differences between the two periods in table 1. Effectively, during the 10 elements the trampolinist performs complex gestures which combine various rotations following the transverse and longitudinal axes. These variations generate variable blur (Figure 3). Motion quantity is more important during the routine and at this point the bounding box is effective.

The setup sequence has been used to develop the system. We denote good results for this sequence. The test sequence is a beginner trampolinist practising a different routine. Results show that the system is efficient throughout the 10 elements.

	setup sequence		**test sequence**	
Body part	**A**	**B**	**A**	**B**
hands	73 %	70 %	72 %	72 %
head	97 %	96 %	94 %	92 %
base	100 %	100 %	97 %	96 %
knees	99 %	100 %	69 %	97 %
feet	85 %	92 %	59 %	87 %

Table 1. Body parts belonging to the bounding box in different parts of the sequence. A : Entire sequence. B : the 10 competition elements.

An image is computed in 30 milliseconds and the video real time constraints show an image every 40 milliseconds. We can theoretically reach video real time. We say theoretically because in our evaluation, we used non-compressed image sequences which are too heavy to perform real time processing at 720 x 576 resolution.

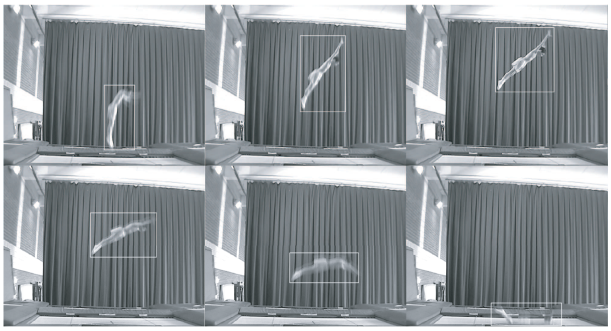

Fig. 4. Bounding box computed by the system on a part of the sequence (every 6 frames). (f 1 0 /)

Figure 4 shows the bounding box created as part of the evaluation sequence. We can see in this figure the heavy blur on the arms. We can adjust the camera shutter speed and open the iris to achieve better image quality (sharpen images).

6 Conclusion and Future Works

We have devised a system which allows efficient tracking of a natural sporting practice. It is not the athlete who adapts to the machine but the machine which adapts to the athlete. Our context is trampoline competitions; consequently the conditions for the trampoline are ideal (no intrusive system). We have an effective characterization of the movements that allows us to design a system which leads to tracking in video real time . The evaluation part denoted good results of bounding box quality and this study emphasizes a significant problem regarding image quality. Acrobatic sports generate fast variable blur on different body parts. Nevertheless, the results are sufficient to develop the next phase. Future work will include many improvements for tracking and acquisition to allow live processing. Then we will implement automatic segmentation of each body part in the bounding box. The information necessary for processing will be considerably reduced because the architecture results in a smaller amount of data for computing. The last phase will include recognition of real time trampolinist motion based on our characterization model.

References

1. T. Molet, R. Boulic, D.: A Real-Time Anatomical Converter for Human Motion Capture. In: Proc. 7h Eurographics Workshop on Animation and SimulationS, Springer-Verlag, Wien (1996)
2. Aggarwal, J., Cai, Q.: Human motion analysis: A review. Computer Vision and Image Understanding: CVIU **73** (1999) 428–440
3. N.R. Howe, M.E. Leventon, W.F.: Bayesian reconstruction of 3-d human motion from single-camera video. Technical Report TR9937-12, ANIPS (1999)
4. Bottino, A., Laurentini, A.: Non-intrusive silhouette based motion capture. In: Proc. of 4^{th} World Multiconference on Systemics, Cybernetics and Informatics, SCI 2000, Orlando (FL) (2000)
5. Lee Campbell, A.B.: Recognition of human body motion using phase space constraints. Fifth International Conference on Computer Vision **309** (1995) 624–630
6. A. Bharatkumar, K. Daigle, M.P.Q.C.J.A.: Lower limb kinematics of human walking with the medial axis transformation. In: Workshop on Motion of Non-Rigid and Articulated Objects, Austin, Texas, USA (1994)
7. M. Yamamoto, T. Kondo, T.Y.K.Y.: Skill Recognition. In: Third IEEE International Conference on Automatic Face and Gesture Recognition. (1998)
8. J. Perš, G. Vučkovič, S. Kovačič, Branko Dežman: A low-cost real-time tracker of live sport events. In: ISPA 2001 Pula, Croatia, University Computing Center, University of Zagreb (2001) 362–365
9. Fédération Internationale de Gymnastique: Code of points. FIG Store, www.fig-gymnastics.com (2001)
10. Leung, M.K., Yang, Y.H.: First Sight: A Human Body Outline Labeling System. In: IEEE Trans. on PAMI. Volume 17. (1995)
11. R. Cassel: Etude sur le suivi et la reconnaissance des mouvements en temps réel d'un trampoliniste par traitement d'images. Master's thesis, École doctorale en informatique de l'Université Paris Sud & LIMSI-CNRS, Orsay (2002)
12. C. Collet, R. Gherbi: Visual perception tools for natural interaction a gaze capture and tracking system. In: IEEE SSIAI'98, Tucson, Arizona, U.S.A (1998) 91–96
13. Haritaoglu, I., Harwood, D., Davis, L.: w^4 s: A real-time system for detecting and tracking people. In: Computer Vision—ECCV. (1998)