

Temporally Synchronising Image Sequences using Motion Kinematics

Kwangwu Lee and Richard D. Green

University of Canterbury, Department of Computer Science.

Email: kwl14@student.canterbury.ac.nz

Email: richard.green@canterbury.ac.nz

Abstract

In this paper, we introduce a method for temporally synchronising image sequences of human body movement. The idea is based on matching similar moving objects in two video sequences to synchronise using temporal features. We describe the experiment with real-time image sequences of human movement skills, being “summersaults” in this case study. Based on two principle motion features—the centre of mass and the principal axis—calculated from movement sequences, we discuss a detailed implementation, measurement of the difference between two moving objects, performance, and limitations of the proposed algorithm. The results indicate that the temporal synchronising algorithm proposed is useful for synchronising the movement of objects in video sequences.

Keywords: computer vision, machine vision, motion analysis, sport coaching, human body, temporal synchronization

1 Introduction

Given two image sequences of an object in similar motion, one of the most basic and useful task would be temporally matching the videos by synchronising the object’s motion. It is even more useful when this is applied to identifying and categorising many patterns of human movement such as walking, running, or jumping. However, implementing this operation which is cognitively easy for humans, on a computer system is a challenging problem. In general sequences of motion, synchronising two similar moving objects, which are varying in their velocities and the directions could be a difficult task, if it has to be done manually by a simple video editing tool. Having a robust method to automatically track objects and synchronise videos based on some invariant features is the goal of this research.

In computer vision literature, modelling the human body and its motion is an area of great interest to many researchers, as survey [1] shows. The research can be roughly grouped into two parts: tracking method with model reconstruction, and movement recognition. Already impressive results have been reported in the recognition area alone. Yamato et al. [2] reported successful recognition of six different tennis strokes among three subjects. Starner and Pentland [3] also worked on hand gesture system, American Sign Language, with a 99% successful recognition rate. These are based on

use of Hidden Markov Models (HMM). Recently, Green and Guan [4] [5], which are also based on HMM, proposed more unified and general framework for recognising human movement patterns.

For tracking the human body, a popular method is CONDENSATION (CONDITIONAL DENSITY PROPAGATION) algorithm [6], or also known as particle filter [7]. Because of multiple predictions for each variables being tracked, it is more robust to heavy background clutter than the traditional Kalman filter [8].

In this paper, we demonstrate the experiment with “salto” as our case study example, which is a somersault with a full 360° rotation about the transverse axis of the body. With video to be synchronised consisting of relatively simple and elegant movement sequences, the salto is an ideal choice for the study of human motion in a controlled indoor environment. This research proposes a computationally efficient temporal synchronising algorithm using the centre of mass and principal axis calculated in real-time from a sequence of images.

2 Background

It is assumed that only one main moving object is captured by a fixed camera environment. This assumption is necessary to applying a Double Difference Algorithm (DDA) [9] to obtain an object motion mask.

The entire region of the moving object (i.e., the gymnast performing a salto) stays within the image boundary, so that the centre of mass can be always computed correctly.

2.1 Overview of the Procedure

An outline of the general procedure of this experiment is:

1. Apply double difference algorithm (DDA) to isolate the moving object.
2. Apply morphological closing operation to obtain smoothed object [10].
3. Calculate the centre of mass and principal axis angle.
4. Apply heuristics to adjust principal axis angle.
5. Calculate the alignment error of two salto actions frame by frame, spatially and temporally aligned on the centre of mass.
6. Calculate the alignment error of two salto actions frame by frame, spatially aligned on the centre of mass and temporally aligned by the principle axis.

2.2 Kinematics of Motion

We exploit the kinematics of saltos by observing that the following two features are the most important parameters: (1) centre of mass and (2) principal axis.

2.2.1 Centre of Mass

According to the law of Newtonian physics, a system of particles can be regarded as a single particle at the position of the centre of mass \mathbf{r}_{cm} with the mass M , which is the sum of all the mass of the system of particles ($M = \sum_{k=1}^{k=n} m_k$). The following equation establishes the position of the centre of the mass by:

$$M\mathbf{r}_{cm} = m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + \dots + m_n\mathbf{r}_n$$

The path of its centre of mass can be replaced as a single particle with the sum of mass at the centre and determined by the initial velocity. As shown figure 2, the green graph line represents the vertical component of the centre of mass position over the horizontal axis which is the timeline based on frame index (33mS/frame @ 30fps). Like any other object thrown into air, the familiar parabolic path of motion is followed using a Physics based model of flight under gravity [11].

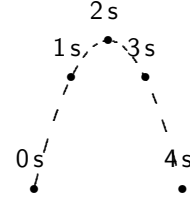


Figure 1: An example of parabolic flight path of the centre of mass.

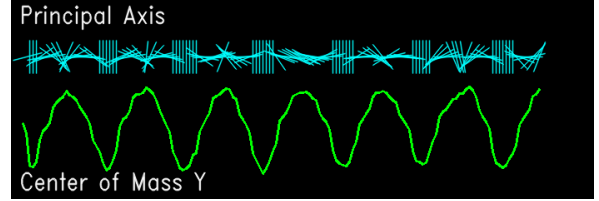


Figure 2: Plot of multiple saltos. Centre of mass Y and principal axis angle is represented on frame index (33mS/frame).

2.2.2 Principal Axis

The angle of the principal axis of an object is based on the rotational inertia and is calculated as follows using a special case equation [12] of 2D binary image:

$$\tan^2 \theta + \frac{\mu_{20} - \mu_{02}}{\mu_{11}} \tan \theta - 1 = 0 \quad (1)$$

where the second order of moments μ_{20} , μ_{11} , and μ_{02} are defined as:

$$\mu_{20} = \sum_{i=1}^m \sum_{j=1}^n B[i, j](x_{ij} - \bar{x})^2 \quad (2)$$

$$\mu_{11} = \sum_{i=1}^m \sum_{j=1}^n B[i, j](x_{ij} - \bar{x})(y_{ij} - \bar{y}) \quad (3)$$

$$\mu_{02} = \sum_{i=1}^m \sum_{j=1}^n B[i, j](y_{ij} - \bar{y})^2 \quad (4)$$

and $B[i, j]$ is intensity value (0 or 1) of an m -by- n binary image at (i, j) , x_{ij} and y_{ij} are the vertical and horizontal coordinates of the image, respectively. \bar{x} and \bar{y} are the centre of the mass of horizontal and vertical spatial position, respectively.

Solving equation 1 yields:

$$\theta = \frac{1}{2} \arctan \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right) \quad (5)$$

Calculating the angle of the principal axis has a number of issues. Unless it is masked out, the motion of the trampoline bed can generate horizontal noise as a part of the moving object causing errors of the angle as high as 90° , even though the body posture on the bed is close to straight upright position.

Another problem is the range of the angle θ ($0^\circ < \theta \leq 180^\circ$) from equation 5, which corresponds to the upper or lower half of the full circle. This causes a 180° ambiguity of upside-down between the two videos. To solve this problem, a constant angular velocity is assumed (or at least the conservation of angular momentum) so any sudden change of angle greater than 90° is assumed to be a shifting to the upper or lower range of circle.

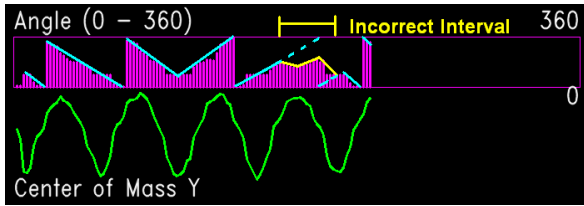


Figure 3: An example of principal axis angular velocity flipping 180° angle. The dotted-line represent the projection of expected angle.

Finally, a tucked body shape in the salto action can cause the minimum and maximum axes to have similar values, which results in a poor estimate of the principal axis. As shown in figure 3, at the fourth peak of parabolic graph, the angle begins to decrease. This eventually loses the tracking angle and can flip to an opposite angle causing an error of 180° . The conservation of angular momentum L can be described as:

$$L = I\omega = \text{a constant}, \quad (6)$$

or

$$I_i\omega_i = I_f\omega_f. \quad (7)$$

where I is the rotational inertia and ω is the angular velocity.

While an object in flight may change the rotational inertia from its initial value I_i to a smaller or larger value I_f , the rotational inertia cannot become negative value. Therefore, while the angular speed can be increased or decreased, the direction of rotation cannot reverse while in flight.

In other words, the sign of the angular velocity (i.e. the tangent of the upper graph in figure 3) cannot be changed, and the angle should be either monotonously increasing or decreasing. The change of angular velocity direction is allowed to

change only at the lowest point of the centre of mass while in contact with the trampoline bed.

This problem is solved by (1) maintaining the sign of angular velocity until next landing, or by (2) calculating the angular momentum, when the rotational momentum becomes as small as a certain threshold value, we interpolate the angle based on current angular velocity.

3 Results

3.1 Performance

Intel OpenCV Library and Microsoft MFC were used on a 2GHz PC with 512MB memory. The 30fps video clips were adjusted below 20 frames per second to avoid dropped frames returned by a slow OpenCV frame grabbing function.

3.2 Experimental Result

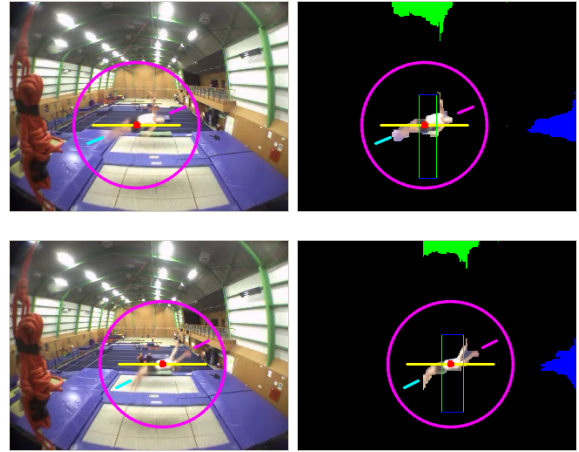


Figure 4: Tracking of centre of mass and principal axis through two video sequences.

Figure 5 shows a number of features of the given frame that may be useful for synchronisation. The blue histogram on right and the green histogram on top is the projection of image onto a line obtained by summing rows of pixels and summing columns of pixels. The yellow horizontal line inside the circle is a visual cue for 0° . From the two line segments in the circle, one can see the principal axis. A projection image histogram is used to identify sudden noise increases such as movement of the trampoline bed.

The box represents the maximum and the minimum centre of mass over all the frames. To compute these values, all frames need to be passed once. During the first pass, the frame indices of the local extremum—the highest and the lowest point—of the centre of mass, are calculated.

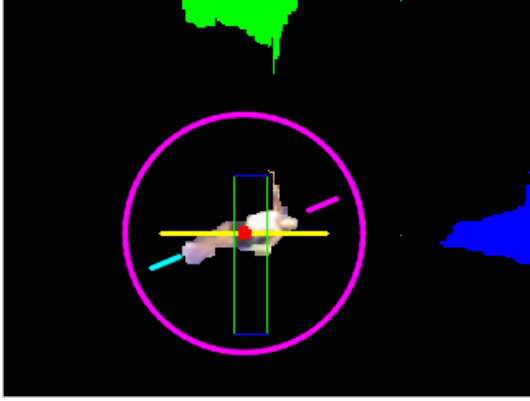


Figure 5: An analytical version including: the maximum and the minimum centre of mass, principal axis, and histograms of projection.

Figure 6 shows the two gymnasts overlapped at the centre of mass. One can easily see the difference of the principal axes.

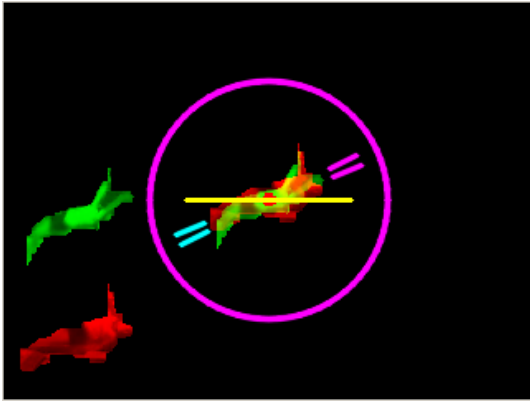


Figure 6: An image with two gymnasts overlapped.

3.3 Measurement of Difference

In this experiment, the measure of how closely the two human body images match is calculated as the ratio of the number of pixels overlapped (intersection) to the number of total number of non-zero intensity valued pixels (union). In figure 7, the last third column images show overlap of the first and the second column images. The first row is overlapped at the centre of mass, while the second row images are also aligned on the principal axis as well. Clearly, in this synchronisation, the overlapped image with both centre of mass and the principal axes aligned together produce better match by 45% than that without alignment the principal axis 30% by 15% improvement. However, in other cases, alignment of principal axis sometimes produced worse results. Surprisingly,

the alignment of the principal axes of the two images do not always produce better synchronisation results.

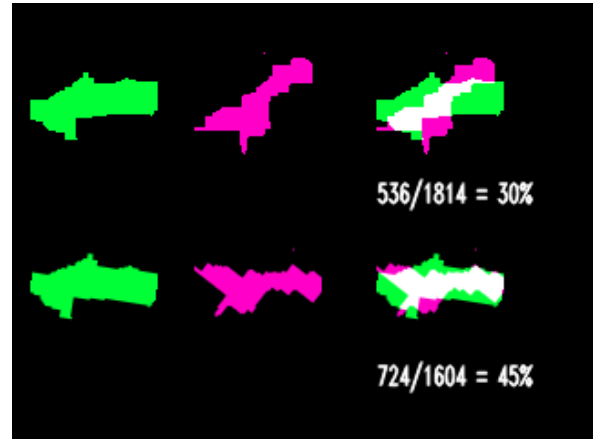


Figure 7: The first row illustrates the alignment error using centre of mass synchronisation. The second row illustrates the alignment error using principal axis synchronisation.

4 Conclusion

The experiment demonstrates the detailed discussion of computing kinematics features of motion in image sequences. The results indicate that the proposed algorithm is a useful method for matching two different motion sequences for comparison and analysis.

We have introduced solutions to the problems of calculating principal axis angles, and a set of novel methods to measure the difference (or distant) of two image sequences. This temporal synchronisation algorithm could also be applied to a number of other applications such as video retrieval.

However, the lack of an explicit human model increases the alignment error since an articulated body model is required for more accurate body alignment.

In future work, this method will be extended to include particle filter [7] based approaches with an articulated human body model [13][14].

References

- [1] T. B. Moeslund and E. Granum, "A Survey of Computer Vision-Based Human Motion Capture," *Computer Vision and Image Understanding: CVIU*, vol. 81, no. 3, pp. 231–268, 2001.
- [2] J. Yamato, J. Ohya, and K. Ishii, "Recognizing Human Action in Time-Sequential Images

- using Hidden Markov Model,” in *Proc. Conf. on Computer Vision and Pattern Recognition*, pp. 379–385, 1992.
- [3] T. Starner and A. Pentland, “Real-Time American Sign Language Recognition From Video Using Hidden Markov Models,” in *SCV95*, p. 5B Systems and Applications, 1995.
- [4] R. D. Green and L. Guan, “Quantifying and Recognizing Human Movement Patterns From Monocular Video Images – Part I: A New Framework for Modeling Human Motion,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 2, pp. 179–190, 2004.
- [5] R. D. Green and L. Guan, “Quantifying and Recognizing Human Movement Patterns From Monocular Video Images – Part II: A New Framework for Modeling Human Motion,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 2, pp. 191–198, 2004.
- [6] M. Isard and A. Blake, “CONDENSATION – conditional density propagation for visual tracking,” *International Journal of Computer Vision*, vol. 29, no. 2, pp. 5–28, 1998.
- [7] J. Deutscher, A. Blake, and I. Reid, “Articulated body motion capture by annealed particle filtering,” in *CVPR*, pp. 126–133, 2000.
- [8] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *Transactions of the ASME—Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.
- [9] J. K. Aggarwal and Q. Cai, “Human Motion Analysis: A Review,” *Computer Vision and Image Understanding: CVIU*, vol. 73, no. 3, pp. 428–440, 1999.
- [10] L. G. Shapiro and G. C. Stockman, *Computer Vision*. Prentice-Hall, 2001.
- [11] D. Halliday, R. Resnick, and J. Walker, *Fundamentals of Physics*. John Wiley & Sons, 4 ed., 1993.
- [12] R. Jain, R. Kasturi, and B. Schunck, *Machine Vision*. McGraw-Hill, 1995.
- [13] R. Plänkers and P. Fua, “Articulated Soft Objects for Video-based Body Modeling,” in *ICCV*, (Vancouver, Canada), July 2001.
- [14] J. M. Rehg and T. Kanade, “Model-Based Tracking of Self-Occluding Articulated Objects,” in *ICCV*, pp. 612–617, 1995.