INTER-TRIAL DIFFERENCE ANALYSIS THROUGH APPEARANCE-BASED MOTION TRACKING

Dan Mikami[1], Toshitaka Kimura[1], Koji Kadota[2,3], Makio Kashino [1,3], and Kunio Kasino[1]

NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, Kanagawa, Japan[1]
Graduate School of Medicine, Osaka University, Osaka, Japan[2]
Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Japan[3]

The purpose of this study is to develop a method for quantitative evaluation and visualization of inter-trial differences in the motion of athletes. Previous methods for kinematic analyses of human movement have required attaching specific equipment to a body segment or can only be used in an environment designed for analyses. Therefore, they are difficult to use for observing motions in real games. To enhance the applicability to real-game situations, we propose appearance-based motion tracking. Our method only requires an image sequence from a camera. From the image sequence, automatic detection of trials and a difference analysis of them are conducted. We applied our method to the analysis of pitching motions in actual baseball games. Though we have no quantitative evaluations yet, the experimental results imply the efficacy of our method.

KEY WORDS: tracking, motion history image, synchronized replay, self-organizing feature map.

INTRODUCTION: Quantitative evaluation of the motions of athletes is important for enhancing performance in sports. Quantitative feedback immediately after games plays an important role in performance improvement. Several studies of athlete kinematics have been performed (Moeslund & Granum, 2001). Basically, image-based motion analysis has mainly focused on entertaining viewers. Therefore, the analysis granularity is too rough for feedback to coaches and athletes. On the other hand, most non-image-based motion analysis methods require attaching specific equipment to a body segment, wearing special clothes, or working in a special environment designed for the analysis. These requirements narrow the coverage of previous kinematic analysis methods, especially in actual game situations and in handling large numbers of trials. We aim at overcoming the problems by devising a novel appearance-based motion tracker.

METHODS: We propose a novel inter-trial difference visualization method that is available for actual game situations. Our method targets motions that are similar to each other and that occur repeatedly. Our method automatically detects target motions from a video. And then, it analyzes small differences among them. One typical example of target motion is the baseball pitching motion.

The key component of our method is our new appearance-based motion tracker. The appearance-based motion tracker employs a motion history image (MHI), as proposed by Bobick & Davis (2001), to represent motions. It consists of two stages: a template registration stage and a tracking stage. In the template registration stage, MHI-based template images are obtained. They are composed of a sequence of template images. In the tracking stage, our method estimates the motion category by retrieving the most similar image among the template images from the MHI created at each time step. At the same time, it also estimates the position that the most similar template MHI is obtained. With our appearance-based motion tracker, each motion can be quantitatively evaluated. Though the target in this study is the baseball pitching motion in real games, our method can be applied to other movements, such as those in tennis, track and field, and gymnastics. Remainder of this section briefly reviews the MHI and then describes each stage.

MHI: The MHI, a method of motion representation proposed by Bobick & Davis, has been widely used because of its ease of implementation. Figure 1 shows an MHI and snapshots of the corresponding image sequence, where the snapshots are shown from left to right in time order. In the MHI, the value of each pixel shows how recently a motion was detected on the pixel. Bright (white) pixels denote pixels at which motions are detected. As the time proceeds from the most recent motion, the pixels turn dark.

Figure 1: MHI (left) and the snapshots of the corresponding image sequence (right). The motion is raising the left leg.

The pixel value of MHI $H(x,y,t)$ at position (x,y) and time t can be obtained by

$$H(x,y,t) = \begin{cases} 255 & D(x,y,t) = 1 \\ H(x,y,t-1) - g & \text{otherwise,} \end{cases} \qquad (1)$$

where 255 (i.e., white) is a pixel value for pixels on which a motion is detected, and $D(x,y,t)$ denotes motion detection function. As the motion detection function, inter-frame difference is commonly used. In addition, g denotes a decay parameter; if small g is employed, the resulting MHI is affected by motions of long past.

Template registration stage: The proposed appearance-based motion tracker is a template-matching-based tracker. Therefore, it requires registering a template as a reference for the template matching in the motion tracking stage. Unlike the previous MHI-based motion detectors, our method uses a time series of MHI-based template images as a template. A start time, an end time, and a target region of template creation are set by user input. Figure 2 shows a time series of MHI templates.

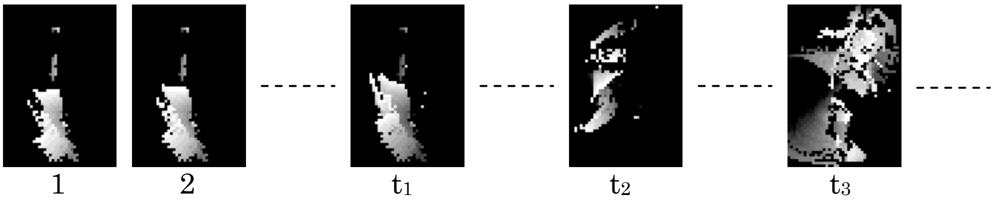

Figure 2: An MHI template sequence. MHIs are listed left to right in time order.

Motion tracking stage: In the motion tracking stage, a motion category and the motion centre position where it is observed are estimated at each time step. The motion category is selected from the registered MHIs. From the temporal transition of the motion category (e.g., Figure 3, the left most image), automatic detection of trials can be done. In addition, the outputs, i.e., motion category and motion centre positions, of detected trials are used for a detailed analysis of motions.
The procedure in motion tracking stage is as follows: (1) Create a MHI for a new input image. (2) Compare the new MHI to registered MHI templates. (3) Determine the most similar MHI template and the position at which it is observed. We use the Euclidean distance to evaluate the similarity.
Figure 3 shows a set of outputs for one trial of pitching. Here, we used a baseball practice match captured by a Sony HandyCam HDR-CX560 with 30 [fps] and 1024 x 768 [pixels] resolution. The left figure shows the temporal development of the determined motion category. The middle and right figures show the transitions of horizontal and vertical positions, respectively. As shown in Figure 3, our appearance-based motion tracker well estimates the motion category and position.

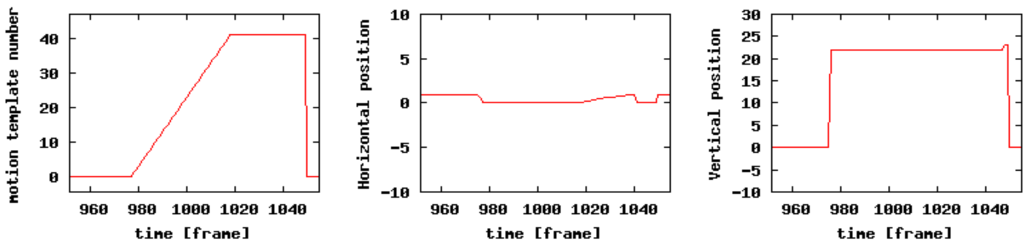

Figure 3: Examples of motion tracking outputs. Left: transition of recognized motion category. Middle and right: transitions of horizontal and vertical positions, respectively.

RESULTS: This section shows two ways to visualize inter-trial differences with our appearance-based motion tracker; one is a synchronized video replay and the other is motion similarity-based mapping.

<u>Synchronized video replay:</u> Quantitative evaluation enables comparison between two trials by means of a synchronized replay. Figures 4 and 5 show snapshots of the application. Each image is composed of a video window (left) and transitions of motion category, horizontal position, and vertical position (right) of detected motion centre. In the left side (video window), two videos are displayed. In the right side, there are three graphs which have yellow and green lines. The yellow lines correspond to the left video in the video window, and green line corresponds to the right video in the video window. The vertical red lines denote current time.
Figures 4 and 5 show the result for the same trial pair, and show different temporal positions. Here, the left video is normal pitching, and the right video is quick motion pitching to prevent steals. At the time in Figure 4, the left normal pitching has already started. On the contrary, the right quick pitching has not started yet. At the time in Figure 5, both motions are almost the same. However, the vertical position of the quick pitching (green line) is lower than that of the normal pitching (yellow line). Currently, we have not conducted a quantitative comparison to motion capture. However, from Figs. 4 and 5, the precise analysis of motion centre points was confirmed.

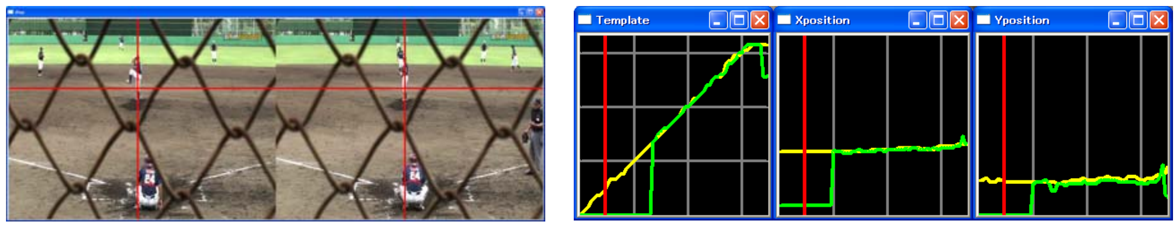

Figure 4: Video window (left) and the transitions of motion category, horizontal position, and vertical position (right). The red line in each transition window denotes current time.

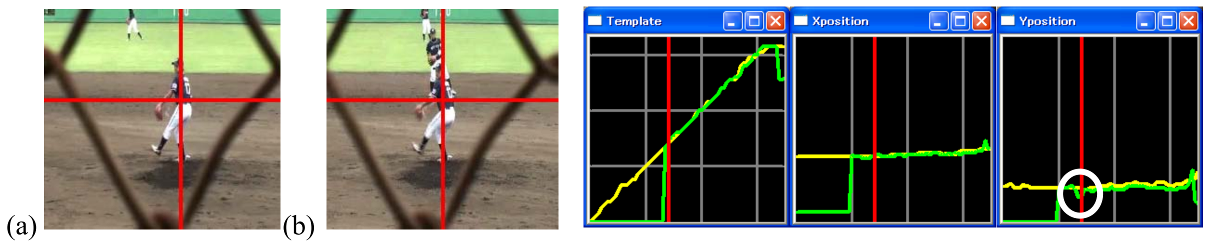

Figure 5: Images (a) and (b) are cropped from the video window. The pitcher sunk more in quick-motion pitching (b) than in normal motion pitching (a). This was successfully measured as a transition of horizontal position (denoted by the white circle).

<u>Motion similarity-based mapping:</u> Kohonen (2001) proposed a self-organizing feature map (SOM). The SOM algorithm projects a set of high-dimensional data into a low-dimensional map keeping their similarity, i.e., similar data pairs are projected into near places in the low-dimensional map. On the contrary, dissimilar pairs are projected into distant places in the low-dimensional map. From the features, the SOM is often used for enhancing visibility of high-dimensional data.

We apply the pitching motion data obtained by our appearance-based motion tracker to the SOM. Each pitching motion is composed by 129-dimensional data. Figure 6 shows a SOM projection result for 39 pitching trials. Trials #1 and #39 are projected into near places in the map (solid circle in Figure 6). Their temporal transitions of motion category, horizontal position, and vertical position are quite similar to each other. On the contrary, Trials #7 and #10 are mapped into distant positions in the map (dashed circle in Figure 6). Their transitions include differences as shown on the right side of Figure 6.

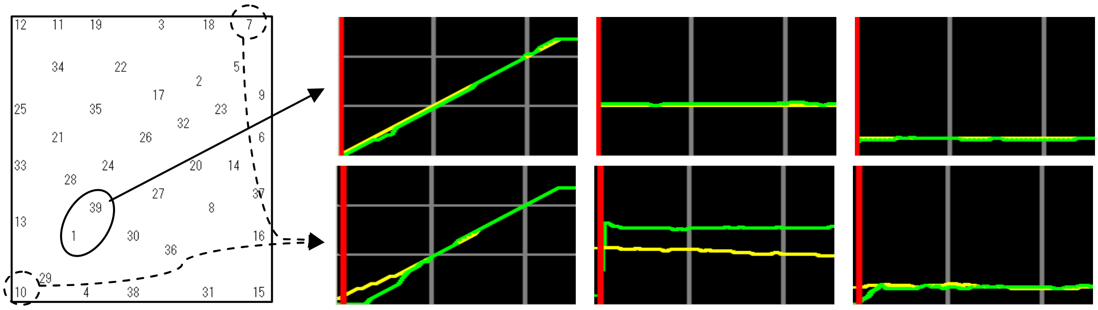

Figure 6: Projected 39 pitching motions and examples of corresponding temporal transitions of motion category, horizontal position, and vertical position.

DISCUSSION: The most important merit of our method is its availability for actual games. This is derived from the requirement to input, i.e., an image sequence captured from a camera. In addition, such an easy capturing environment enables our method to analyse large numbers of trials. In motion capture systems, analysing large numbers of trials is difficult due to severe constraints for capturing. We consider that analysing a large number of trials will tell us much information. As an example, we think that motion variance among trials may become a good index of proficiency.

Currently, our method only detects comprehensive motion information; it is not able to detect detailed, e.g., kinematic, motion differences. We plan to devise our method for enabling more detailed motion differences. In addition, we plan to conduct a correspondence analysis of our method and motion capture and enhancing our method to for detailed motion analyses.

CONCLUSION: We proposed an appearance-based motion tracking method for quantitative visualization of differences in motions. We believe that this novel technique usefully provides comprehensive information and feedback for more natural movements in sports, though it does not have enough resolution to detect detailed differences in motion at present.

REFERENCES:

Moeslund, T. B. & Granum, E. (2001). A Survey of Computer Vision-Based Human Motion Capture, *Computer Vision and Image Understanding*, 81(3), 231-268.

Bobick, A. F. & David, J. W. (2001). The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3), 257-267.

Kohonen, T. (2001) *Self-organizing maps*. Springer.