Human Motion Analysis under Actual Sports Game Situations
Sequential Multi-decay Motion History Image Matching

Dan Mikami[1], Toshitaka Kimura[2], Koji Kadota[3], Harumi Kawamura[1] and Akira Kojima[1]

[1]*Media Intelligence Laboratories, NTT, 1-1 Hikarino-oka, Yokosuka, Kanagawa, Japan*
[2]*Communication Science Laboratories, NTT, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa, Japan*
[3]*Graduate School of Medicine, Osaka University, 1-17, Machikaneyamacho, Toyonaka, Osaka, Japan*

Keywords: Motion Analysis, Motion History Image, MHI, Multiple Decay Parameter, Baseball.

Abstract: This paper proposes a sequential multi-decay motion history image matching with the aim of analyzing human motions captured in actual game situations without subjecting people to any intrusive measures. The motion history image (MHI) is a well- known motion representation method, which can be used without foreground detection. In MHIs, pixels on which motion is detected have large pixel values. As time elapses following the latest motion detection, the values decrease according to a decay parameter. Two improvements were made to enable MHI-based template matching to be applied to motion analysis; introducing a template MHI sequence matching process that enables analysis of the temporal development of motions and extending MHIs to include multiple decay parameters. Due to the MHI sequence, a reference motion includes target motions of various speeds. Since the appropriate decay parameter varies with motion speed, no one predefined decay parameter can be the best one. These improvements enable our method to effectively analyze human motions in actual game situations. Experiments carried out indoors with capturing of 3D motion data and outdoors under real games situations verified the effectiveness of the proposed method.

1 INTRODUCTION

Human motion analysis is one of the most important research areas in the field of computer vision. Its widespred applicability ranges from automatic surveillance and human-computer interaction to biomechanics and rehabilitation. Human motion analyses for automatic surveillance and/or human-computer interaction (Mikami et al., 2009) require recognition of motion categories independent from persons. In other words, a given human motion analysis needs to absorb the person-dependent motion differences and to recognize the motion category.

On the other hand, in cases when a human motion analysis aims at a quantification of motions for biomechanical and/or rehabilitations purposes, slight differences among multiple trials of the same motion become significant information (Vasconcelos and Tavares, 2008).

The target of this paper is analysis of repetitive human motion; the proposed method aims at analysis and visualization of small differences among trials.

Conventionally, human motion analyses for sports biomechanics have used motion capture systems. Al-

though these systems can effectively acquire 3D position information of body parts, they have severe capturing limitations. These limitations are as follows:

1. Equipping of markers

   Though there are some marker-less motion capture systems, most commercially available systems require that target persons to be equipped with markers to enable their movements to be observed. In addition, to make the markers visible, target persons are required to wear a body-fitting cloth.

2. Illumination conditions

   Motion capture systems are basically designed to be used in laboratories and do not work in direct sunlight.

3. Calibration of multiple cameras

   Motion captures require multiple cameras that need to be calibrated. For example, the well-known commercial motion capture system "Qualisys" requires at least three cameras for motion capturing. In addition, once a camera is moved, the calibration process needs to be carried out again.

Some previous studies have made use of depth sensors such as "Microsoft Kinect" for analyzing human motions (Oikonomidis et al., 2011; Shotton et al., 2011). These sensors also have limitations, however, on items such as sensor-target distance and illumination conditions. As a result, analyzing the motions of athletes in actual game situations is a still challenging problem.

We aim at developing a human motion analysis method that is completely non-intrusive, i.e., requiring neither special device nor body-fitting cloth, thus making it suitable for use in actual game situations.

The motion history image (MHI) approach, which was proposed by Bobick and Davis (Bobick and Davis, 1996; Bobick and Davis, 2001), is acknowledged as a motion analysis and representation method that is robust against capturing environments. Each pixel value of an MHI represents a temporal distance from the latest motion detected at the pixel. Bright pixels denote pixels in which motions are detected, and with the elapse of time following the most recent motions, the pixels become dark. As a result, the MHI resembles an afterimage. The degree of to which pixels become dark is controlled by a decay parameter.

A lot of MHI-based motion representation and detection studies have been carried out. For example, gradient information is used for enhancing sensitivity of both pose and directional motion information (Bradski and Davis, 2002), motion history volumes, which is an extension of the input from 2D image to 3D volume data, was proposed as a free-viewpoint motion representation (Valstar et al., 2004), and multilevel intervals for MHI creation was proposed to overcome self-occlusion problem (Weinland et al., 2006). The most important advantage of the MHI approach is its robustness under various capturing environments. In addition, MHI-based motion detection can be applied to an image sequence without any calibrations.

In the context of motion detection in sports, Mikami et al. used MHI for detecting pitching scenes from baseball videos. In (Mikami et al., 2007), a reference pitching motion is represented by an MHI, and then pitching motions in the target video are retrieved by the reference motion. This method detects pitching motions with high accuracy. However, it is not able to analyze the temporal development of motions.

To the best of our knowledge, temporal development of motion is not targeted by MHI-based motion analysis. This paper proposes a sequential multi-decay MHI matching process that includes two important improvements over existing MHI template matching approaches. First, the proposed method newly introduces a temporal sequence of MHIs to represent a reference motion. By comparing a reference MHI sequence with MHIs from the target video, it simultaneously detects and analyzes the motion. Its use of sequential reference MHIs enables to analyze differences in temporal development among the motions.

Second, the method extends existing MHI to include multiple decay parameters. This compensates for the innate problem of sequential matching. The reference motion sequence includes both quick and slow motions. A small decay parameter for quick motion yields an MHI with many bright pixels, which deteriorates the spatial resolution of analysis. On the other hand, a large decay parameter for slow motion may yield an MHI with no or only a few motion history, which also deteriorates detection accuracy. Consequently, no one predefined decay parameter can be the best one. If the MHI-based method is to be extended to include sequential MHI matching, it must be able to handle variations in motion speed.

In this paper, we use pitching motions in a baseball game as the target of analysis. Our method can be more widely applied, however, to analyzing repetitive motions such as tennis serves and golf swings.

The remainder of this paper is organized as follows. Section 2 reviews the MHI method. Section 3 proposes a temporal MHI sequence matching process. Section 4 shows experimental results and Section 5 concludes the paper with a summary of key points.

2 MOTION HISTORY IMAGE: MHI

The MHI approach, a method of motion representation proposed by Bobick and Davis (Bobick and Davis, 1996; Bobick and Davis, 2001), has been widely used because of its ease of implementation. Many studies to enhance the method have been carried out, as well as many studies using MHI as a motion representation methods have been carried out. Since these have been well described in the literature (Ahad et al., 2012), we will introduce only the basic idea and implementation of the MHI, here.

Figure 1 shows an MHI and snapshots of the corresponding image sequence 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. With the elapse of time following the most recent motion, the pixels turn dark.

The pixel value of MHI, $H(x,y,t)$ at position (x,y)

(a) MHI   (b) Corresponding image sequence

Figure 1: (a) The MHI. (b) Snapshots of the subject raising his leg. As shown in (a), a moving foreground can be obtained without foreground detection.
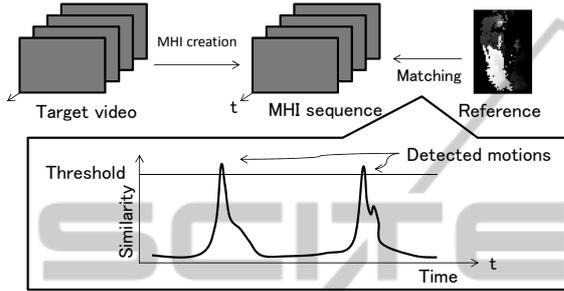

Figure 2: Conventional MHI-based motion detection. At each time step, an MHI is created from the original image sequence. Target motions are detected on the basis of comparison with the reference and subsequent thresholding of similarity.

and time t can be obtained by

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

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

Template matching-based motion detection methods that use MHI for motion representation have been proposed. Similarity criteria have also been proposed. The simplest criterion in these methods is an inverse of Euclidian distance as shown by

$$S(i,j) = \frac{1}{\sum_{x,y} H_i(x,y) - H_j(x,y)}. \quad (2)$$

If the similarity is larger than a threshold, the target motions are detected (Fig. 2).

MHI-based motion detections, the decay parameter g is predetermined on the basis of motion speed. The most appropriate decay parameter varies with motion speed. If a too small decay parameter is used for fast movements, many motions are mixed together and the precise detection of a target motion becomes difficult. On the other hand, if too large decay parameter is used for slow movements, the ability of motion

expression becomes low. Therefore, accuracy of motion detection will be deteriorated.

3 PROPOSED METHOD: SEQUENTIAL MULTI-DECAY MOTION HISTORY IMAGE MATCHING FOR MOTION ANALYSIS

The proposed method detects and analyzes repetitive human motions simultaneously by comparing them with a reference motion. In our method, a reference motion is represented by a sequence of MHIs.

To enable MHI-based template matching to be applied to motion analysis, the proposed method improves the existing MHI-based template matching procedure in two ways:

1. Expanding a template MHI to a temporal sequence of MHIs to represent a reference motion (sequential MHI matching),

2. Expanding an MHI to a set of MHIs with multiple decay parameters (multi-decay MHI matching).

The former enables analysis of the temporal development of a motion. The latter is necessary to obtain good matching between sequences. This is an important improvement because a template, being a set of temporal sequences by the former improvement, may include both fast and slow movements. And, most appropriate decay parameter varies on the basis of motion speed.

The proposed method consists of two steps; a template registration step and a motion detection/analysis step. At the motion registration step, the area and start and end time of reference motion are manually set. It generates the sequential multi-decay MHI. Then, at the detection/analysis step, the most similar MHI among the sequential multi-decay MHI, similarity between them, and a position where it was found were obtained for each time step.

The improvements are described in the next two subsections.

3.1 Sequential MHI Matching

The proposed method uses a sequence of MHIs to represent a reference motion. Hereafter, we refer to it as the "reference MHI sequence." Figure 3, a template ID is applied to each image in the reference MHI sequence; the ID corresponds to the amount of time (in frames) from the beginning of the template motion.
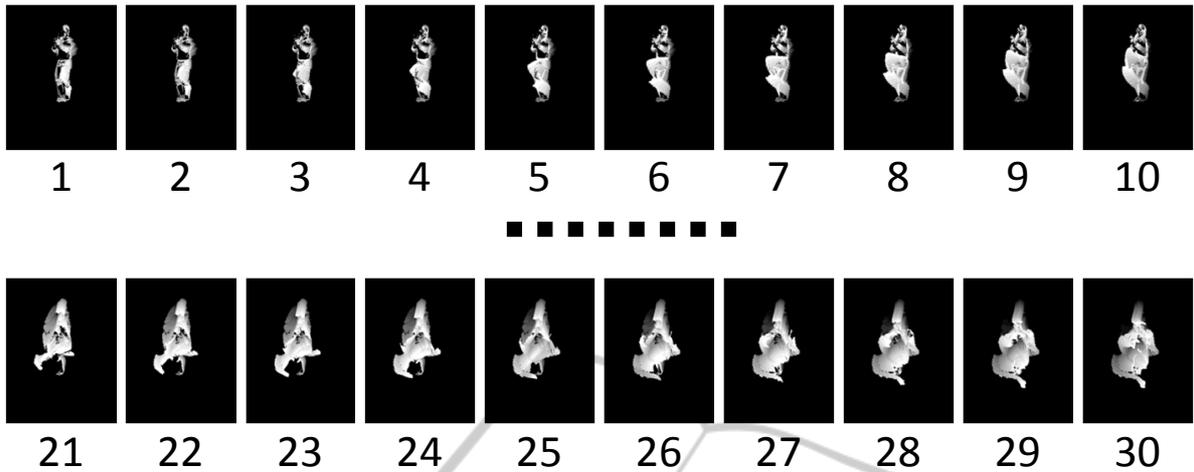
Figure 3: Example sequence of MHIs; a template ID is assigned to each MHI. To be exact, each MHI is extended to multi-decay MHI as described in Section 3.2.
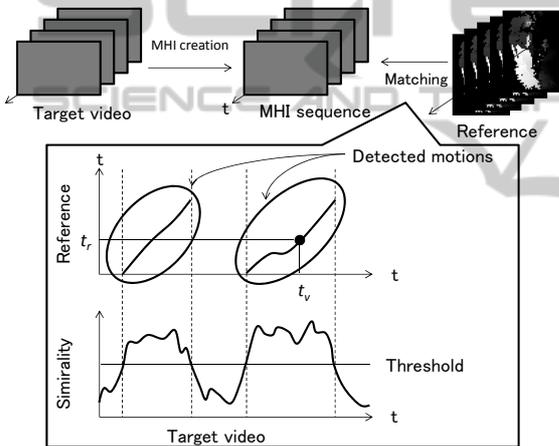

Figure 4: Proposed MHI-based motion detection. At each time step, an MHI is created from the original image sequence. Target motions are detected on the basis of comparison with the reference and subsequent thresholding of similarity.

The proposed method uses the reference MHI sequence $\boldsymbol{R} = \{R(1), \cdots, R(L)\}$ to retrieve the motion from the MHI sequence created from a target video, where L is the number of images within the reference. At each time step, it obtains the template ID of the most similar of the MHIs in the reference MHI sequence. At the same time, the similarity of the MHI with the retrieved most similar template MHI and the position where it is retrieved are also obtained. The graph at the top of balloon in Fig.4 denotes the temporal transition of the retrieved template ID. The dot in the graph demonstrates that the MHI whose template ID $= t_r$ is the closest to the MHI at time t_v. The graph at the bottom shows the transition of the similarity between them.

A lot of sequence matching methods have been proposed; dynamic time warping (DTW) and Hidden Markov Model are the most famous examples and they can be applied to our method. However, to simply verify the effectiveness of using the reference MHI sequence, the proposed method first detects the most similar of the MHIs in the reference MHI sequence. Then, on the basis of the temporal transition of the template ID, it simultaneously detects and analyzes the target motion.

Our sequential MHI matching approach can be written as follows:

$$\widetilde{k}(t) = \operatorname*{argmin}_{k \in \{1, \cdots, L\}} S(H(k), R(k)), \qquad (3)$$

$$m(t) = \sum_{i=1}^{L} |i - \widetilde{k}(t+i)|. \qquad (4)$$

If $m(t)$ is lower than a threshold, the proposed method detects it as a targeted motion.

3.2 Multi-decay MHI

Figure 5 shows MHIs of the decay parameters 4, 8, 16, 32, and 64 for the same motion. The motion is that of a pitcher pitching as seen from a side view; the pitcher raises his left leg, steps forward, and throws the ball. In Fig. 5, the horizontal axis is time. The top row shows the original images and the lower rows show MHIs with decay parameters in increasing order.

A small decay parameter for a quick target's movement yields an MHI that includes too much past motion information; template matching using such an MHI deteriorates spatial resolution of motion detection. In contrast, a large decay parameter for a slow
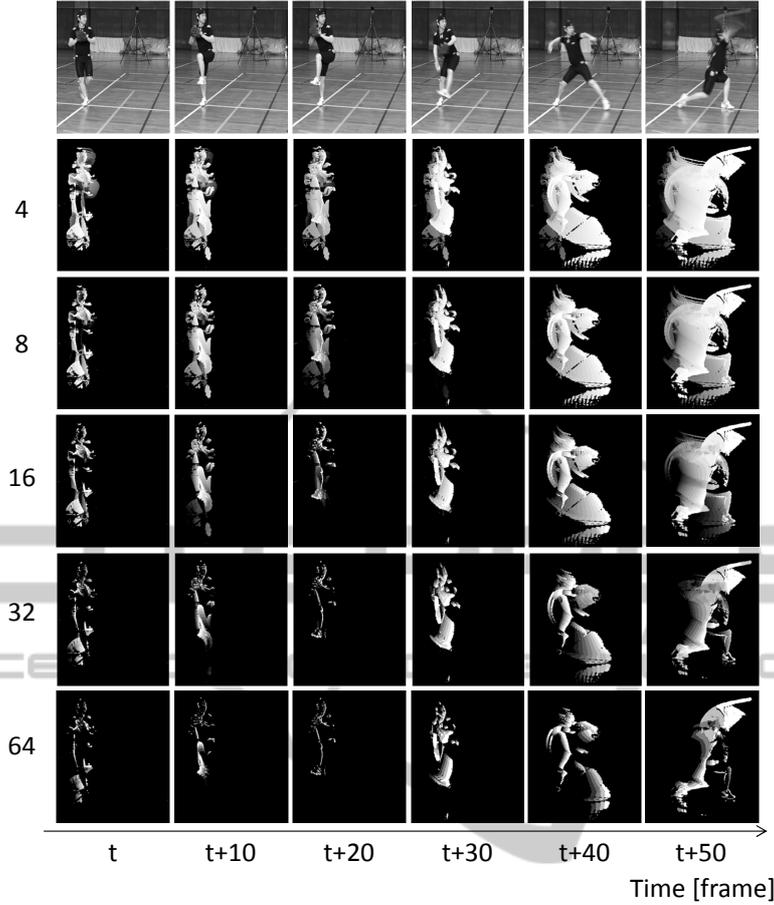
Figure 5: MHIs of different decay parameters. The lines show MHIs with the decay parameters 4, 8, 16, 32, and 64, respectively. The left row shows the view immediately after the pitching motion started.

movement yields an MHI that includes no or only a few motion information; such an MHI cannot provide a good motion template.

3.2.1 Definition of Multi-decay MHI

As can be observed in Fig. 5, appropriate decay parameters are affected by motion speed. However, since the matching method has been expanded to MHI sequence matching as described in Sect. 3.1, the reference motion includes both quick and slow movements. Therefore, no decay parameter is unable to yield good MHIs for the sequence. To overcome this problem, we expand the MHI to a new multi-decay MHI, which is actually a set of MHIs with multiple decay parameters.

The new multi-decay MHI, $M(t)$, at time t is defined as

$$M(t) = \{H^{(1)}(t), \cdots, H^{(g)}(t), \cdots, H^{(G)}(t)\}, \quad (5)$$

where $H^{(g)}(t)$ denotes MHI at time t with decay parameter g. It can be re-written as follows.

$$M(x,y,t) = \{H^{(1)}(x,y,t), \cdots, H^{(g)}(x,y,t), \\ \cdots, H^{(G)}(x,y,t)\}, \quad (6)$$

where $M(x,y,t)$ denotes a set of pixel values at (x,y), which is a G dimensional vector. Here,

$$H^{(g)}(x,y,t) = \begin{cases} 255 & D(x,y,t) = 1, \\ H(x,y,t-1) - g & otherwise, \end{cases} \quad (7)$$

the same as in previous MHI.

3.2.2 Similarity between Multi-decay MHIs

The similarity between multi-decay MHIs $\Phi(M(i),M(j))$, i.e., the similarity between $M(i)$ and $M(j)$, is defined as follows.

$$\Phi(M(i),M(j)) = S(H^{(\widehat{g})}(i), H^{(\widehat{g})}(j)) \quad (8)$$
$$\widehat{g} = \underset{g}{\operatorname{argmin}} |var(H^{(g)}(i)) - V|, (9)$$

where $var(H^{(g)}(i))$ denotes a variance of pixel values within $H^{(g)}(i)$. This means that the decay level
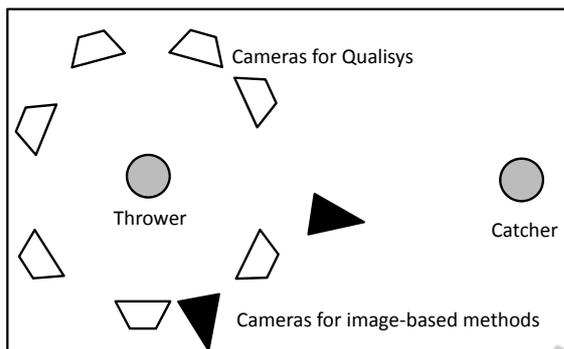
Figure 6: Experimental setup. Seven cameras were used for Qualisys (motion capture system). Two cameras were used for capturing videos.

$\widehat{g}$ is selected from the variance of $H^{(g)}(i)$, and then the MHIs with decay level $\widehat{g}$ are used for similarity calculation. Note that, decay level $\widehat{g}$ employed for comparison target is also $\widehat{g}$. Here, $\Phi(M(i), M(j))$ is a pseudo-distance, so $\Phi(M(i), M(j)) \neq \Phi(M(j), M(i))$. However, it is not a problem for retrieving similar movements. The variable V is a parameter to determine appropriate variance, which is determined as 50 experimentally.

4 EXPERIMENTAL

To verify the effectiveness of our method, we conducted experiments by using two types of videos of pitchers throwing a baseball. Those of the first type were in a gym and the motion capture system "Qualisys" was used for taking measurements. With this system we captured 3D positions for 28 of pitcher's joints. Two cameras were used simultaneously to take the videos; the settings are shown in Fig. 6. A total of 121 pitching trials were captured using two subjects.

Those of the second type were taken during actual baseball games; pitching motions made during the game were detected and analyzed.

The sequential template MHIs were set manually.

4.1 Motion Analysis with Proposed Method

4.1.1 Effect of Sequential Matching

Figure 7 shows an example output of our proposed method and Fig. 8 shows the vertical position obtained from 3D motion camera data for corresponding trials.

In Fig. 7, the reference motion is shown in the top left window and the detected motion is shown in the
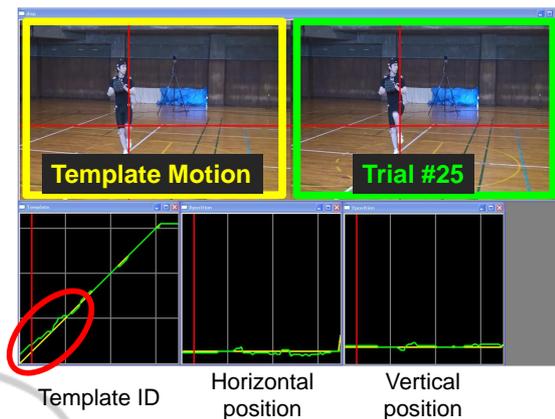

Figure 7: Results obtained with proposed method. The top left window shows the video of the template motion; the top right window shows a video of a detected trial. Both motions are synchronized. The bottom row shows temporal transitions of the detected template ID (left) and those of positions where the templates are detected (middle and right).

top right one. For purpose of visibility, normal videos are shown, although MHIs are used for detection and analysis. The detected template ID and the transitions of detected positions are depicted at the bottom.

As can be seen from the red circled area in Fig. 7, templates with IDs larger than that of the reference motion are detected at the same timings at the detected motions. This means that the detected motion starts faster than the template motion. After that, both motions are synchronized well.

This was also observed from motion capture data (Fig. 8). Checking the movements of the left toe showed that the detected movement (green line) started faster than the reference movement (red line). Good synchronization of both movements was also observed. In contrast, no differences were observed for the right hand and the right toe.

As described here, our proposed method comprehensively obtains movements and visualizes differences between motions. This intuitive output is one of the method's most attractive advantages. Motion capture data makes it possible to obtain precise three-dimensional data of body joints. Although it makes detailed analysis possible, there is a possibility that important information may be hidden within such voluminous data.

4.1.2 Effect of Multi-decay MHI

To verify the effectiveness of introducing multi-decay MHI, we compared the proposed method to sequential MHI matching with fixed decay parameters. The results are shown in Fig. 9, which shows temporal
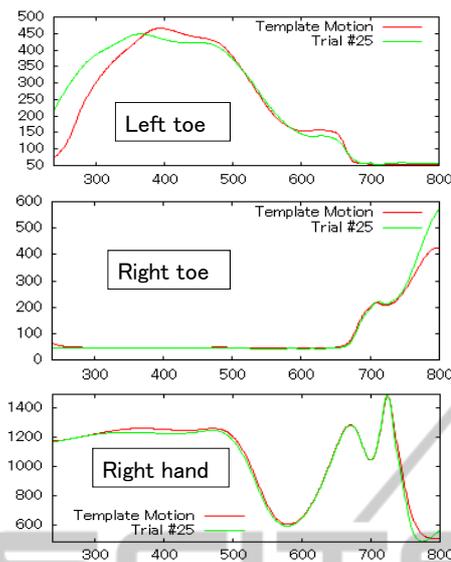
Figure 8: The 3D position data obtained via motion capture. The target motion is the same as that in Fig. 7. The graphs show the elevation occurring within the 3D data of the joints.

(a) ID

(b) Horizontal position

(c) Vertical position

Figure 9: Temporal transition of ID that is the most similar and positions where it is found.

transition of the detected ID and positions where it is observed. As shown in Fig. 9, when the fixed decay parameter 64 was used, the matching was failed from frame 1105 to frame 1110. On the other hand, cases that the decay parameter was 16 and the multi-decay parameter, the correct matching were done.

Figure 10 shows the MHI at frame 1107 and the MHIs of frame 25 and 85 from the beginning of the template sequence; frame 25 of template MHI sequence was selected as the most similar one for frame 1107. The MHIs with decay parameter 64 included quite few motion information. As a result of low motion representation ability, the matching failed. On the contrary, the MHIs with decay parameter 16 include rather much motion information, which led the correct matching of motions.

4.2 Other Examples

Figure 11 shows another example of motion detection and analysis by our method. On the basis of the transitions of matched template ID, detected motion, which is denoted by green, started faster than reference, which is denoted by yellow. This can be well observed from the snapshot.

During the pitching motion, horizontal position of detected motion became large as shown at the bottom graph. Though the difference is quite small and is difficult to be recognized, our method well visualized such small difference.
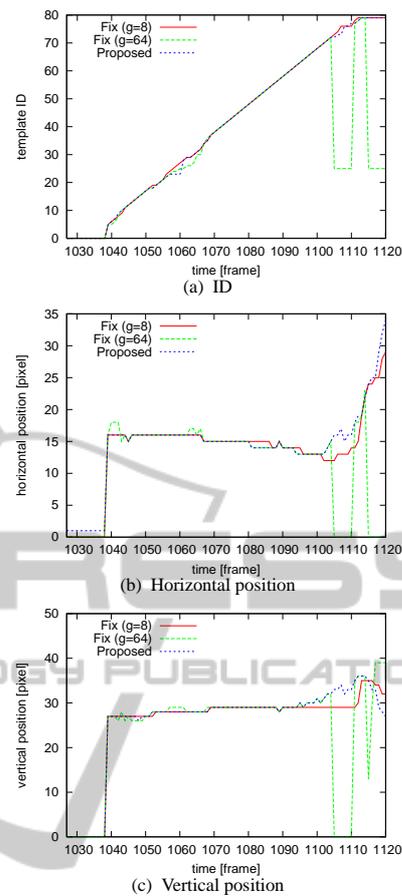
5 CONCLUSIONS

This paper described a sequential multi-decay motion history image (MHI) matching method we have developed and here propose with the aim of analyzing human motions made in game situations without subjecting subjects to any intrusive measures. Two improvements were made to enable MHI-based template matching to be applied to motion analysis. The first is introducing a template MHI sequence matching process and the second is extending MHIs to include multiple parameters. These improvements enable our method to effectively analyze human motions in actual game situations.

Future work will include developing an analysis method to improve the association between results and body parts. At present our method handles movements comprehensively; however, more detailed analysis should is required to improve its performance.
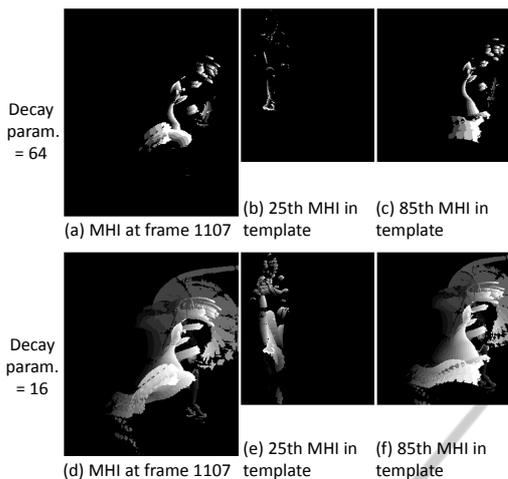
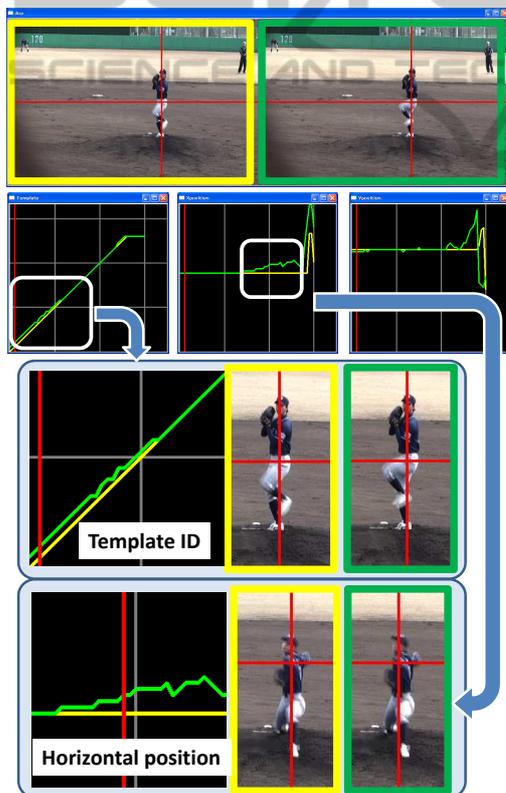Figure 10: Comparison of created MHI at frame 1107 and multi-decay MHIs in the template.

Figure 11: Another example of motion detection and analysis.

REFERENCES

Ahad, M. A. R., Tan, J. K., Kim, H., and Ishikawa, S. (2012). Motion history image: its variants and appli-
cations. *Machine Vision and Applications*, 23(2):255 – 281.

Bobick, A. F. and Davis, J. W. (1996). Real-time recognition of activity using temporal templates. In *Proc. of 3rd IEEE workshop on Applications of Computer Vision (WACV)*, pages 39 – 42.

Bobick, A. F. and Davis, J. W. (2001). The recognition of human movement using temporal templates. *IEEE Trans. PAMI*, 23(3):257 – 267.

Bradski, G. R. and Davis, J. W. (2002). Motion segmentation and pose recognition with motion history gradients. *Machine Vision and Applications*, 13(3):174 – 184.

Mikami, D., Konya, S., and Morimoto, M. (2007). Pitch by pitch event detection using impulse sound detection and moving image clustering. *IEICE Trans. Information and Systems*, J90-D(2):526–534.

Mikami, D., Otsuka, K., and Yamato, J. (2009). Memory-based particle filter for face pose tracking robust under complex dynamics. In *Proc. of IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 999 – 1006.

Oikonomidis, I., Kyriazis, N., and Argyros, A. A. (2011). Efficient model-based 3D tracking of hand articulations using Kinect. In *Proc. of BMVC*.

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., and Blake, A. (2011). Real-time human pose recognition in parts from a single depth image. In *Proc. of IEEE Computer Vision and Pattern Recognition (CVPR)*.

Valstar, M., Pantic, M., and Patras, I. (2004). Motion history for facial action detection in video. In *Proc. of IEEE Int'l conf. on System, Man and Cybernetics*, volume 1, pages 635–640.

Vasconcelos, M. J. M. and Tavares, J. M. R. (2008). Human motion analysis: Methodologies and applications. In *Proc. of Int'l Symposium on Computer Methods in Biomechanics and Biomedical Engineering*.

Weinland, D., Ronfard, R., and Boyer, E. (2006). Free view point action recognition using motion history volumes. *Computer Vsion and Image Understanding*, 104(2–3):249–257.