The Thirteenth International Symposium on Artificial Life and Robotics 2008(AROB 13th '08),
B-Con Plaza, Beppu, Oita, Japan, January 31-February 2, 2008

Human Motion Representation and Recognition by Directional Motion History Images

Masayuki Fukumoto, Takehito Ogata, Joo Kooi Tan, Hyoung Seop Kim, Seiji Ishikawa

Department of Control Engineering, Kyushu Institute of Technology, Kitakyushu, Japan

(E-mail: {fukumoto,etheltan,ishikawa}@ss10.cntl.kyutech.ac.jp)

Abstract: In this paper, we propose a technique for recognizing human motions using directional motion history images. A motion history image represents a human motion as a single image produced by superposing binarized motion image frames so that older frames may have smaller weights. It has, however, difficulty that a newer motion is overwritten on older motions, resulting in inexact motion recognition. In order to overcome this difficulty, we propose directional motion history mages which describes a motion by separating it to four directions of movement, i.e., right, left, up and down, employing optical flow and representing each of them by a motion history image. Experimental results indicate that the proposed technique shows better performance in the recognition of rather complicated human motions than simple motions which could be recognized by the existent motion history images.

Keywords: Motion recognition, motion history image, optical flow.

I. INTRODUCTION

In recent years, studies on automatic recognition of human motions have drawn much attention among computer vision researchers supported by high performance of computer. Recognition of human motions can be applied to various fields that need human surveillance; monitoring a suspicious person in front of a cash dispenser, watching senior people or patients for preventing them from accidents, and so on.

One of the techniques on human motion recognition is based on human posture recognition [1]. However this technique has to extract a human area from a background. Actually it is not very simple to extract a human area in a complicated background and the precision of the extraction directly influences the performance of the recognition.

Bobick et al. [2] proposed a method of human motion description using a motion history image (MHI). It is applied not only to motion recognition [2,3], but also to other fields [4,5,6]. A MHI is generated from binarized difference images between successive image frames representing a single motion and superposing them so that the older image frames may have smaller weights. Binary difference images are easy to be calculated even from an image sequence having complicated backgrounds. However the application of MHI to complex motions often causes trouble, because older motion is overwritten by the present motion, resulting in incomplete motion description and therefore it may cause inexact motion recognition.

To overcome this difficulty, we propose a technique for motion recognition employing directional motion history images (DMHI) [3] that represents four direction of motions, i.e., up, down, right and left, by individual MHIs. This paper claims better performance of DMHI than MHI, when they are applied to the recognition of complex motions.

II. GENERATION OF DMHI

A MHI is one of the methods for representing time-varying images that leaves past images on a single image. When a MHI is generated from time-varying images, newer images are presented brighter, whereas older images darker. Let us define binary images $D(x,y,t)$ (t=1,2,…,T) by

$$D(x,y,t)=\begin{cases}1 & \dots\ |F(x,y,t)-B(x,y)|>\theta \\ 0 & \dots\ otherwise\end{cases}, \tag{1}$$

where $F(x,y,t)$ (t=1,2,…,T) is the original video image sequence and $B(x,y)$ is the background. Then a MHI is defined by

$$H_\tau(x,y,t)=\begin{cases}\tau & \dots\ D(x,y,t)=1 \\ \max(0,H_\tau(x,y,t-1)-1 & \dots\ otherwise\end{cases}, \tag{2}$$

where τ is an integer signifying previous τ image frames are taken into account.

On the other hand, the DMHIs that we propose are generated from optical flow images of the original by resolving them into four directions, i.e., up, down, right and left. Let us denote the optical flow at pixel (x,y) and time t on image frame F by $v(x,y,t)\equiv v$. Then v is decomposed into four directions and their intensities are stored at pixel (x,y). The procedure is described as follows;

$$(x,y)\leftarrow \left(\max\{0,i^{\mathrm{T}}v\},\min\{|i^{\mathrm{T}}v|,0\},\max\{0,j^{\mathrm{T}}v\},\min\{|j^{\mathrm{T}}v|,0\}\right) \tag{3}$$

Here i and j are unit vectors on the x-axis and on the y-axis, respectively. Transpose of a vector is denoted by 'T'.

Eq.(3) is computed at every sample time t (t=2,3,…,T).

The four optical flow images $F^{x+}, F^{x-}, F^{y+}, F^{y-}$ at time t (t=2,3,…,T) are then defined as

$$
\begin{aligned}
F^{x+}(x,y,t) &= \max\{0, i^{\mathrm{T}} v(x,y,t)\}, \\
F^{x-}(x,y,t) &= \min\{|\, i^{\mathrm{T}} v(x,y,t)|, 0\}, \\
F^{y+}(x,y,t) &= \max\{0, j^{\mathrm{T}} v(x,y,t)\}, \\
F^{y-}(x,y,t) &= \min\{|\, j^{\mathrm{T}} v(x,y,t)|, 0\}.
\end{aligned} \tag{4}
$$

Employing Eq.(4), DMHIs are defined as follows;

$$
\begin{aligned}
H_\tau^{x+}(x,y,t) &= \begin{cases} \tau & \cdots \ F^{x+}(x,y,t) \ge \theta \\ \max\{0, H_\tau^{x+}(x,y,t-1)-1\} & \cdots \ \mathit{otherwise} \end{cases} \\
H_\tau^{x-}(x,y,t) &= \begin{cases} \tau & \cdots \ F^{x-}(x,y,t) \ge \theta' \\ \max\{0, H_\tau^{x-}(x,y,t-1)-1\} & \cdots \ \mathit{otherwise} \end{cases} \\
H_\tau^{y+}(x,y,t) &= \begin{cases} \tau & \cdots \ F^{y+}(x,y,t) \ge \theta \\ \max\{0, H_\tau^{y+}(x,y,t-1)-1\} & \cdots \ \mathit{otherwise} \end{cases} \\
H_\tau^{y-}(x,y,t) &= \begin{cases} \tau & \cdots \ F^{y-}(x,y,t) \ge \theta \\ \max\{0, H_\tau^{y-}(x,y,t-1)-1\} & \cdots \ \mathit{otherwise} \end{cases}
\end{aligned} \tag{5}
$$

Here the following is assumed (*=x+, x-, y+, y-);

$$\forall x, \forall y,\ H_\tau^{*}(x,y,0) = 0. \tag{6}$$

Parameter θ' has a role of creating binary input images as in Eq. (1).

Finally we have the following four DMHIs;

$$H_\tau^{x+}(x,y,T), H_\tau^{x-}(x,y,T), H_\tau^{y+}(x,y,T), H_\tau^{y-}(x,y,T), \tag{7}$$

which are a MHI toward x+ direction, a MHI toward x-direction, a MHI toward y+ direction, and a MHI toward y- direction, respectively.

In the same way as Bobick et al. [2], we define directional motion energy images (DMEIs) from DMHIs as follows;

$$
\begin{aligned}
E_\tau^{x+}(x,y,T) &= \begin{cases} 1 & \cdots \ H_\tau^{x+}(x,y,T) > 0 \\ 0 & \cdots \ \mathit{otherwise} \end{cases} \\
E_\tau^{x-}(x,y,T) &= \begin{cases} 1 & \cdots \ H_\tau^{x-}(x,y,T) > 0 \\ 0 & \cdots \ \mathit{otherwise} \end{cases} \\
E_\tau^{y+}(x,y,T) &= \begin{cases} 1 & \cdots \ H_\tau^{y+}(x,y,T) > 0 \\ 0 & \cdots \ \mathit{otherwise} \end{cases} \\
E_\tau^{y-}(x,y,T) &= \begin{cases} 1 & \cdots \ H_\tau^{y-}(x,y,T) > 0 \\ 0 & \cdots \ \mathit{otherwise} \end{cases}
\end{aligned} \tag{8}
$$

A DMEI is a binary image showing the range of the motion in an image. Both DMHIs and DMEIs are employed for motion recognition in the present technique. Actually it is reasonable to consider that, if one of the DMEIs provides a large motion area compared with other three DMEIs, its partner DMHI plays an important role in the recognition, since the DMHI may represent a principal motion among the four DMHIs.

Examples of MHIs and DMHIs are given in **Fig. 1**.

The overwriting issue concerning MHI is solved by DMHIs. In the original motion shown in Fig. 1(a), a subject stoops (a1-a2) and stretches (a2-a3). The MHIs are produced from (a1-a2) as shown in (b1) and from (a1-a3) as given in (b2). Obviously the stoop motion is overwritten by the stretch motion. On the other hand, the proposed DMHIs are depicted in (c), in which motion (a1-a2) is represented by the image H_τ^{y-} in the second row of the left column and motion (a2-a3) by the image H_τ^{y+} in the first row of the right column. In this way, the two opposite sub-motions in the original motion in (a) are clearly separated in DMHIs.

It is noted that optical flows are extracted at moving part of an image and no flow appears on the static background. Hence human mobile region extraction is rather simple in the present technique by use of the optical flow.

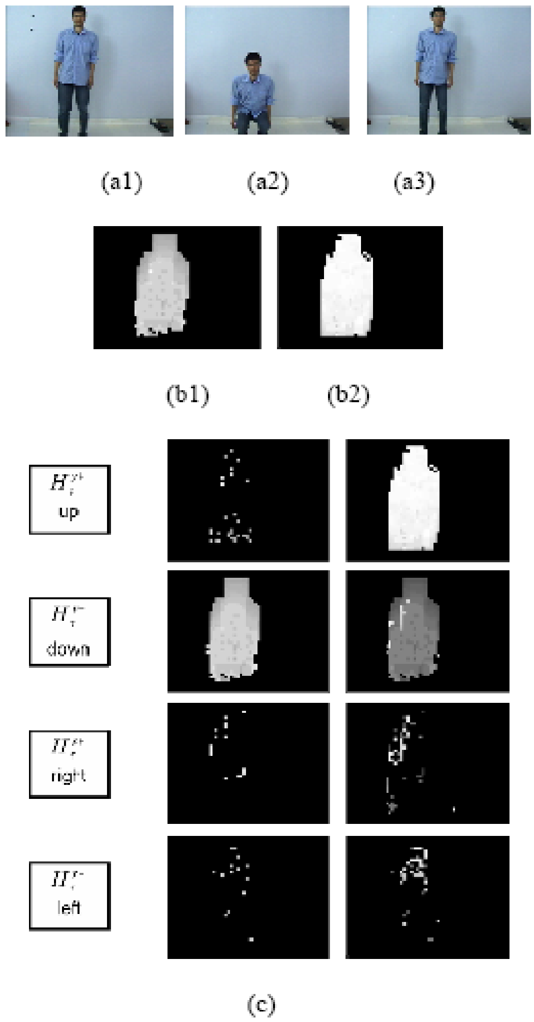

Fig.1 Avoidance of overwriting: (a) Original stooping and stretching motion; (b1) A MHI w.r.t. (a1-a2), and (b2) a MHI w.r.t. (a1-a3); and (c) the proposed DMHIs. As shown in (b2), the MHI in (b1), the stooping motion, is overwritten by the stretching motion, whereas they are clearly separated in (c).

III. MOTION RECOGNITION SCHEME

In order to define a feature space, we choose the 1st to the 7th order Hu moments [7] calculated from the generated DMHIs and DMEIs. Hu moments are calculated from the normalized central moment and have the property that they are invariant to translation, scale and rotation. In order to take account of the volume/area of a motion provided by DMHI/DMEI, the 0th order moment is also considered.

Let the Hu moments calculated from one of the DMHIs be denoted by $\phi_1^H, \phi_2^H, \ldots, \phi_7^H$ and those from one of the DMEIs by $\phi_1^E, \phi_2^E, \ldots, \phi_7^E$. The normalized 0th order moment of the DMHI denoted by $\hat{\mu}_{00}^{H*}$ is defined employing the 0th order moments μ_{00}^{H*} (*=x+, x-, y+, y-) as

$$\hat{\mu}_{00}^{H*} = \frac{\mu_{00}^{H*}}{\mu_{00}^{Hx+} + \mu_{00}^{Hx-} + \mu_{00}^{Hy+} + \mu_{00}^{Hy-}}. \tag{9}$$

In the same way, the normalized 0th order moment of the DMEI denoted by $\hat{\mu}_{00}^{E*}$ is defined by

$$\hat{\mu}_{00}^{E*} = \frac{\mu_{00}^{E*}}{\mu_{00}^{Ex+} + \mu_{00}^{Ex-} + \mu_{00}^{Ey+} + \mu_{00}^{Ey-}}. \tag{10}$$

Then we define the following feature vector with respect to motion m;

$$x_m^* = \left(\phi_1^H, \phi_2^H, \ldots, \phi_7^H, \hat{\mu}_{00}^H, \phi_1^E, \phi_2^E, \ldots, \phi_7^E, \hat{\mu}_{00}^E\right)_m^*, \tag{11}$$

where * represents x+, x-, y+ and y-. This is a feature vector having 16 components. These four feature vectors are collected into a single feature vector having 64 components as follows;

$$x_m = \left(x_m^{x+}, x_m^{x-}, x_m^{y+}, x_m^{y-}\right). \tag{12}$$

This is registered in a 64-dimensional feature space. Note that x_m^* and x_m are column vectors.

In the performed experiment, a motion m is acted by multiple subjects and taken video images by multiple cameras. All these video images are, as learning data, transformed into the feature vectors of the form (12) and registered in the 64-dimensional feature space. Hence a motion m is represented by multiple feature vectors $x_{m,i}$ (i=1,2,…,I_m) provided from different persons and camera orientations.

An observed unknown motion m is similarly transformed into a feature vector x of the form (12). In order to find the closest x_m with x in the feature space, the k nearest neighbor (k-NN) method is employed. The following set S_M is calculated;

$$S_M = \arg k \min_{m,i} \left\{\left\| x - x_{m,i} \right\|\right\}. \tag{13}$$

where, for $a_1 < a_2 < \cdots < a_K$,

$$k \min\{a_i \mid i = 1,2,\ldots,K\} \equiv \{a_1, a_2, \ldots, a_k\} \quad (k<K),$$
$$\arg\{a_1, a_2, \ldots, a_k\} \equiv \{1,2,\ldots,k\}.$$

Finally, the majority motion in S_M is assigned to the unknown motion m.

IV. EXPERIMENTAL RESULTS

Recognition of tennis strokes were conducted in order to show performance of the proposed technique.

Three cameras were set around a subject: 12 subjects acted 5 tennis strokes, i.e., forehand stroke, backhand stroke, forehand slice stroke, backhand slice stroke and smash (See **Fig.2**).

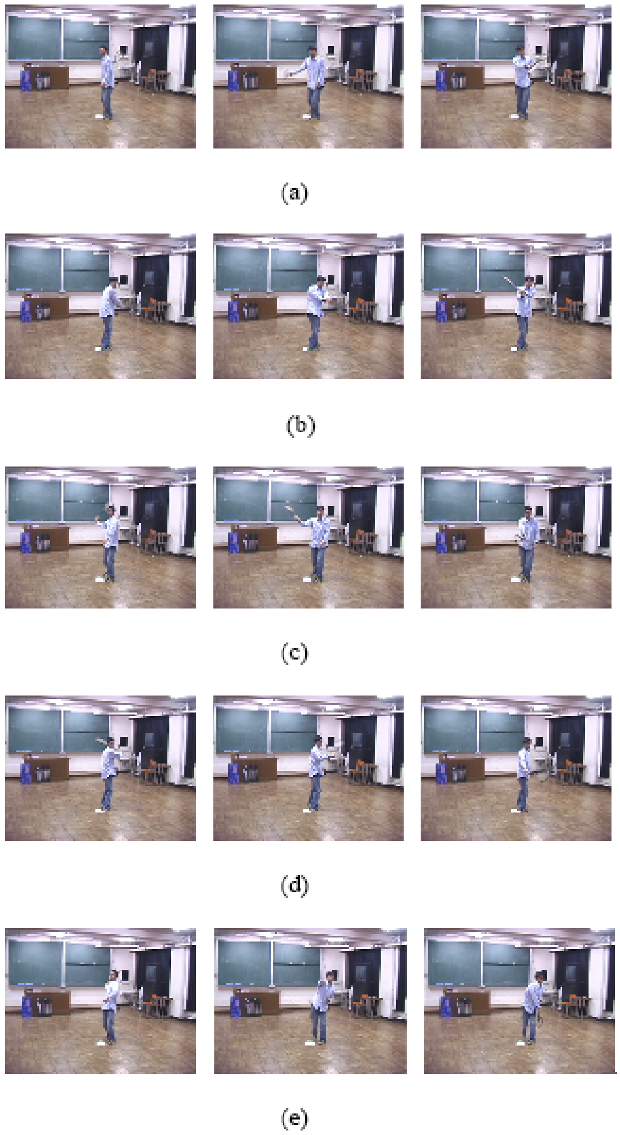

Fig.2 Five tennis strokes in the experiment; (a) a forehand stroke; (b) a backhand stroke; (c) a forehand slice stroke; (d) a backhand slice stroke; and (e) smash.

Table 1 Experimental results: Recognition rates of tennis strokes employing DMHIs and DMEIs compared to that employing only DMHIs.

	DMHI*+DMEI**	DMHI
Forehand stroke	83.3 %	75.0 %
Backhand stroke	75.0 %	75.0 %
Forehand slice	75.0 %	66.7 %
Backhand slice	91.7 %	66.7 %
Smash	75.0 %	100.0 %
Average	**80.0 %**	**76.7 %**

*: Directional motion history images
**: Directional motion energy images

Leave-one-out method was introduced to perform the recognition: 3*5=15 clusters were defined in the 64-dimensional feature space from the training data provided by 11 subjects and the remaining single subject's test data was examined for recognition. This was repeated 12 times and the average recognition rate was calculated. The set parameters are as follows: The parameter τ that decided the length of the past history for DMHI was set at the number of frames of each motion; the threshold θ was 1.0; and with the k-NN method, $k = 3$.

In order to show the effect of DMEI, two experiments were performed. In the first experiment, DMHIs and DMEIs were considered, whereas only DMHIs were taken into account in the second. The result is shown in **Table 1**. The average recognition rate over all the motions was 80.0% in the first case. It reduced to 76.7% in the second though.

V. CONCLUSION

We proposed a new human motion recognition technique employing directional motion history images and successfully recognized 5 tennis strokes. The technique can reduce the difficulty of overwriting issue unavoidable in the MHI method to a certain extent. It is likely that the proposed technique is effective, in particular, to the recognition of relatively complex motions. But, of course, the overwriting problem may occur again even by the proposed technique, if a motion sequence is longer. Basic motions should be defined in the first place. Then a long motion sequence could be segmented to those basic motions by employing the proposed technique. This is now under investigation.

The experimental results indicate effectiveness of the employment of DMEIs. The effect of DMEI may be that it contributes to making the inter-class variance larger and inner-class variance smaller.

We employed optical flow images for making DMHIs referring to the idea of motion descriptor [8]. The process of computing optical flows needs further improvement, since pixel correspondence is sometimes not very simple to find when the chosen tennis strokes are played fast.

If the proposed technique is to be implemented on a mobile intelligent robot in future for recognizing a human motion, the recognition needs to be conducted indifferent to the orientation of observation by the robot. This can be realized by the employment of the motion database proposed by Tan et al.[9].

The authors are grateful to those who played several tennis strokes in the experiment.

REFERENCES

[1] Watanabe T, Yachida M(1998), Real time gesture recognition using eigenspace from multi input image. IEICE Technical Report on Electronic Information Displays 97(527):1-8

[2] Bobick A, Davis J(2001), The representation and recognition of action using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(3):257-267

[3] Ogata T, Tan JK, Ishikawa S(2007), Human motion recognition based on directional motion history images. Proceedings of the International Workshop on Advanced Image Technology 857-862

[4] Valstar M, Pantic M, Patras I(2004), Motion history for facial action detection in video. IEEE Int. Conf. on Systems, Man and Cybernetics 1:635-640

[5] Yin Z, Collins R(2006), Moving object localization in thermal imagery by forward-backward MHI. Proc. Conf. on Computer Vision and Pattern Recognition Workshop 133-140

[6] Kumar S, Kumar D, Sharma A, McLachlan N(2003), Classification of hand movements using motion templates and geometrical based moments. Proc. of ICISIP Int. Conf. on Intelligent Sensing and Information Processing 299-304

[7] Hu M(1962), Visual pattern recognition by moment invariants. IRE Transactions on Information Theory 8(2):179-187

[8] Efros AA, Berg AC, Mori G, Malik J(2003), Recognizing action at a distance. Proc. of Int. Conf. on Computer Vision 726-733

[9] Tan JK, Kouno K, Ishikawa S, Kim HS, Shinomiya T(2007), High speed human motion recognition employing a motion database. Imaging & Visual Computing 36(5):738-746. (in Japanese)