Motion Recognition by Higher Order Local Auto Correlation Features of Motion History Images

Kenji WATANABE[1] and Takio KURITA[2]

[1] *Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba*

[2] *National Institute of Advanced Industrial Science and Technology (AIST)*

[1] *kenji-watanabe@aist.go.jp*, [2] *takio-kurita@aist.go.jp*

Abstract

This paper proposes new features for motion recognition. Higher order local autocorrelation (HLAC) features are extracted from the Motion History Images (MHI). Since MHI calculated from the video images include important motion information, it is expected that HLAC features extracted from MHI have good properties for motion recognition. The proposed features were tested using image sequences of pitching in the baseball games. At first the pitchers were identified from the pitching motions by comparing the sequences of HLAC features using DP matching. The pitchers were recognized 100 % correctly when the image size was 90×90 pixels. Then whether there was the runner on a base or not was identified. The recognition rate of the runners from the pitching motions was 96.7 % when the image resolution was set to 25×25 pixels.

1. Introduction

Recently interest in video surveillance has been rapidly increasing because surveillance cameras are installed at many places for security purposes. Automatic motion recognition or motion analysis makes the detection of unusual motions occurred in front of the camera possible.

For motion recognition or motion analysis, it is important to extract good features from a given image sequence. For gesture recognition, Campbell et al. [1] utilized 3-D data gathered from stereo cameras. Mase [2] used optical flow in the context of facial expression recognition. T.Kurita et al. [3,4,5] used higher order local autocorrelation (HLAC) features [6,7] extracted from PARCOR images for gesture recognition. Kobayashi et al. [8] extended HLAC features to address three-way data analysis and proposed new feature called cubic higher order local auto correlation (CHLAC). They were applied for action and person identification. HLAC features of PARCOR images and CHLAC features have several preferable properties for motion recognition such as shift-invariance, additivity, and robustness to noise in image sequences.

CHLAC features were applied by Nanri et al. [1] for abnormality detection. A linear subspace of CHLAC features extracted from normal image sequences was constructed by Principal Component Analysis (PCA). Then abnormal motion was detected by using the distance from the subspace of the normal motions. Since CHLAC features have additivity and all the normal movements are included in the subspace even for scenes containing multiple person's moves, only abnormal movements can be easily detected by simply measuring the distance.

In this paper, we propose new features for motion recognition. We extract higher order local autocorrelation (HLAC) features from the Motion History Images (MHI) [10,11]. We call them HLAC-MHI features. Since MHI calculated from the video images include important motion information, it is expected that HLAC-MHI features have good properties for motion recognition. Like HLAC features extracted from PARCOR images or CHLAC features, HLAC-MHI features also have preferable properties for motion recognition such as shift-invariance, additivity, and robustness to noise in image sequences. HLAC-MHI features are regarded as simplification of CHLAC features. The number of features of HLAC-MHI is 35 and this is smaller than one of CHLAC features. It is expected that HLAC-MHI features are better than HLAC features extracted from PARCOR images because MHI are calculated directly from image sequence than PARCOR images which are calculated by fitting autoregressive model to each pixel.

To confirm the goodness of the proposed features, the proposed features were tested using image sequences of pitching in the baseball games. At first the pitchers were identified from the pitching motions by comparing the sequences of HLAC-MHI features using DP matching. The pitchers were recognized 100 % correctly when the image size was 90×90 pixels. Then whether there was the runner on a base or not was identified. The recognition rate of the runners from the pitching motions was 96.7 % when the image resolution was set to 25×25 pixels.

2. Method

For motion recognition or motion analysis, we have to extract spatial-temporal information from a given image sequence. In the proposed method, the spatial-temporal information is extracted in two stages. At first, Motion History Images (MHI) are calculated from the image sequence. MHI contain temporal information in the image sequence. Then spatial information is extracted from the MHI by higher order local auto correlation (HLAC) features. The number of HLAC features is 35. It is expected that the extracted HLAC features contain spatial-temporal information in the image sequence. By extracting HLAC features form each of MHI, we can obtain a sequence of feature vectors from the image sequence. To reduce the dimensionality of feature vectors, we can apply Principal Component Analysis (PAC).

To cope with non-uniform changes in the speed of motion, we use Dynamic Time Warping (DTW) or DP matching to compare the sequences of feature vectors.

2.1. Motion history images

Motion History Image (MHI) is a scalar-valued image where pixel intensity is a function of the motion history [10]. The pixel intensity becomes brighter when the pixel moves more recently. It will gradually decay unless new movement at that pixel occurs. The pixel intensity of MHI is defined as follows;

$$H_r(x,y,t) = \begin{cases} \tau & D(x,y,t)=1 \\ max(0, H_r(x,y,t-1)-1) & otherwise \end{cases}, \quad (1)$$

where $H_r(x,y,t)$ is the intensity of pixel (x,y) in the frame t. The duration time of MHI is denoted by τ and it represents the holding times of motion history in the image. $D(x,y,t)$ is a binary image sequence indicating regions of motion. In the following experiments, they are calculated by differencing of successive images. If the difference between two pixel intensities is larger than a threshold, $D(x,y,t)$ is set to 1.

2.2. Higher order local autocorrelation features

To make the features invariant to shift of target in the image frame, we extract higher order local autocorrelation (HLAC) features. It is well known that autocorrelation function is shift invariant. Its extension to higher orders was proposed in [11]. HLAC features are subset of the higher order autocorrelations [6]. It is known that HLAC features are very effective for many applications in the image recognition. The N th-order autocorrelation functions with N displacements $(a_1, \cdots, a_N)$ from the reference point r are defined as follows;

$$x(a_1, \cdots, a_N) = \int I(r) I(r+a_1) \cdots I(r+a_N) dr, \quad (2)$$

where $I(r)$ is the pixel intensity of MHI at the reference point r. The number of these autocorrelation functions obtained by the combination of the displacements over the image is enormous. For practical application, they are reduced to define HLAC features. At first, we restrict the order N up to the second. Then, we also restrict the range of displacements to within a local 3×3 window, because the correlation within local region is much higher than the correlation between far points. By eliminating the displacements which are equivalent by the shift, the numbers of patterns of the displacements are reduced to 35. Fig. 1 shows the 35 mask patterns of HLAC feature, where the symbol "$*$" represents "don't care".

Since MHI contain temporal information in the image sequence and HLAC features can extract the spatial information, HLAC features extracted from MHI contain spatial-temporal information on motion in the image sequence. We call them HLAC-MHI features. The number of HLAC-MHI features is 35 and they are represented as a HLAC-MHI feature vector.

By extracting a HLAC-MHI feature vector from each of MHI, we can obtain a sequence of HLAC-MHI feature vectors from the image sequence.

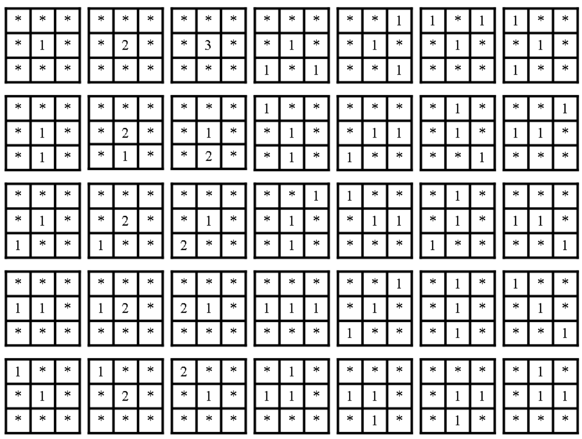

Figure 1. Local mask patterns for computing HLAC features.

2.3. DP matching of sequences of HLAC-MHI feature vectors

For motion recognition we have to define a similarity measure between two sequences of HLAC-MHI feature vectors. Since the speed of motion probably changes in each action, we have to cope with non-uniform changes in the speed of motion. Here we use Dynamic Time Warping (DTW) or DP matching to compare two sequences of HLAC-MHI feature vectors. DTW or DP matching is an algorithm to measure similarity between two sequences of feature vectors and is often used in audio, video, and graphics applications. It finds an optimal match between two sequences by using Dynamic programming. The sequences are warped nonlinearly in the time dimension to determine a measure of their similarity.

If we have a set of training sequences of feature vectors with class labels, then we can classify new sequence of feature vectors by using *K*-nearest neighbor classifier based on DP matching.

In the following experiment we applied Principal Component Analysis (PCA) to reduce the dimensionality of feature vectors.

3. Experiments

To confirm the effectiveness of the proposed HLAC-MHI features, experiments of motion recognition were performed using image sequences of pitching in the baseball games. At first the pitchers were identified from the pitching motions by comparing the sequences of HLAC features using DP matching. Then whether there was the runner on a base or not was identified. We think that these motion recognition tasks are difficult because the differences of pitching motions are very small.

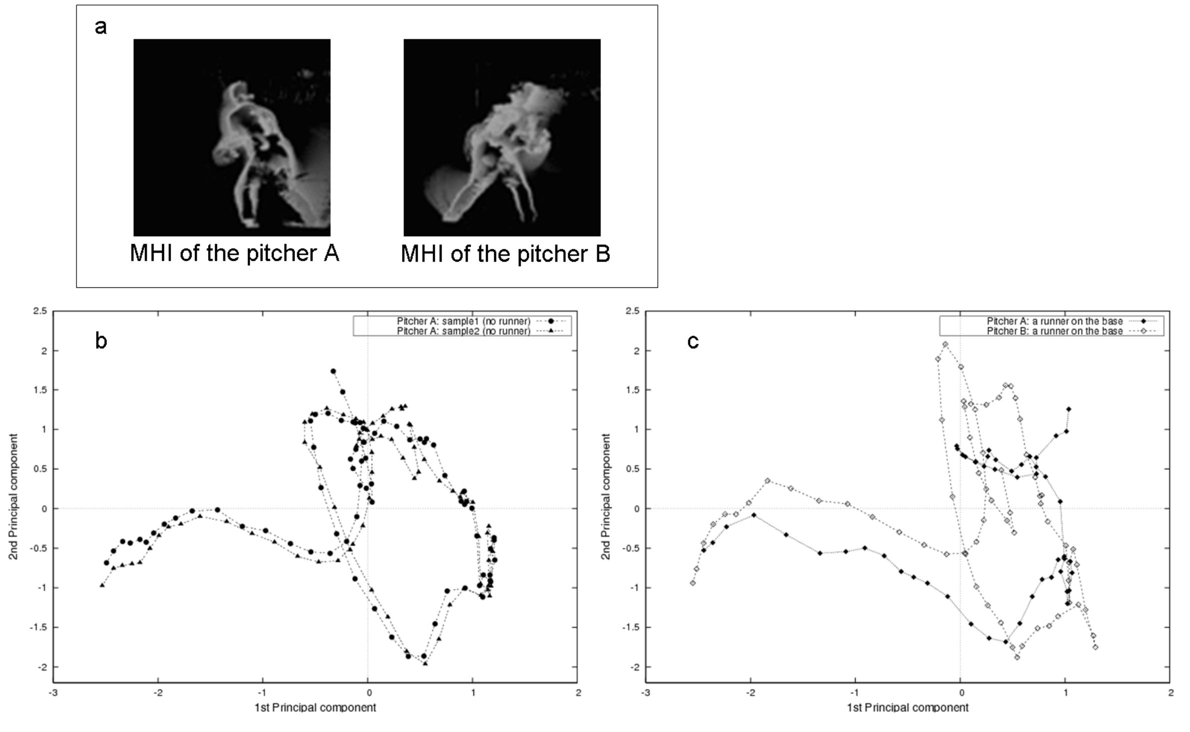

Figure 2. The MHI and the sequences of HLAC-MHI features in the two dimensional PCA subspace. (a) The MHI were calculated from the movies of pitching by the pitcher A and B. (b) The sequences of HLAC-MHI features in PCA subspace by the pitcher A in case there is no runner. Closed circles and triangles respectively denote the HLAC-MHI features of different sequences. (c) The sequences of HLAC-MHI features in PCA subspace for the pitcher A and B in case there is a runner on the base. Closed and open diamonds denote the HLAC-MHI features of the pitcher A and the pitcher B, respectively.

3.1. Pitcher Recognition

Our first experiment is to identify the pitcher using sequences of HLAC-MHI feature vectors. This pitcher recognition task is difficult because the difference of pitching forms of each pitcher is very small. We gathered 30 samples of pitching movies by capturing TV program of a baseball game. They include pitching motions of two pitchers A and B. The size of the movies is 300×300 pixels. These samples were divided to 5 movies of each pitcher (2 samples without runner and 3 samples with a runner on a base) for training and 10 movies of each pitcher (3 samples without runner and 7 samples with a runner on a base) for test.

MHI were constructed from each of these samples. The duration time was set to 1000 ms. Fig. 2 (a) shows the MHI of the pitchers A and B. Then the size of the MHI was reduced from 300×300 to 90×90 pixels and HLAC features were extracted from each of the resized MHI. The dimensionality of the extracted HLAC-MHI feature vectors was reduced by using PCA. The cumulative contribution ratio up to 2nd principal component was over 99 %. The sequences of HLAC-MHI feature vectors in PCA subspace with the 1st and 2nd principal components are shown in Fig. 2 (b) and (c). The sequences of HLAC-MHI feature vectors in Fig. 2 (b) are obtained from the same pitcher A without runner on the first base. The sequences in Fig. 2 (c) are obtained from different pitchers with runner on the first base.

From Fig. 2 (b), the sequences of HLAC-MHI feature vectors in PCA subspace of the same pitcher are very similar. On the other hand, the sequences of HLAC-MHI feature vectors in PCA subspace of different pitchers are the different as shown in Fig. 2 (c). Also it is noticed that the sequences of HLAC-MHI feature vectors were perturbed depending on whether there is a runner or not by comparing Fig. 2 (b) and (c).

To identify the pitchers A and B, we applied *K*-nearest neighbor classifier based on DP matching to the test samples. In this classification experiment, 2 dimensional feature vectors constructed by PCA were used. The recognition rate of the pitcher A and B was 100 %. This result shows effectiveness of the HLAC-MHI features. However, the recognition rate was drooped to 65% when 4 classes (pitcher A without runner, pitcher A with runner on a base, pitcher B without runner, and pitcher B with runner on a base) were classified.

3.2. Runner Recognition

The second experiment is to identify whether there was the runner on a base or not from the movements of pitchers. This task is more difficult than the pitcher recognition task. Similarly we gathered 40 samples of the pitching movies of the pitcher B. The pitching movies were sampled from the different games. These samples are randomly divided to 10 training samples (6 samples without runner and 4 samples with a runner on the first base) and 30 test samples (25 samples without runner and 5 samples with a runner).

MHI were constructed from each of these samples. Then HLAC features were extracted from each of the resized MHI. The dimensionality of the extracted HLAC-MHI feature vectors was reduced by using PCA. *K*-nearest neighbor classifier based on DP matching was used for classification. In this experiment we evaluated the recognition rates by changing the resolution of MHI and the duration time.

Fig. 3 shows the recognition rates when the resolutions of MHI are changed from 10×10 to 300×300 pixels. Here the duration time was set to 1000 ms. The recognition rate is 83.3 % at the 90×90 pixels which is the same resolution with the previous experiment on pitcher recognition. The minimum recognition rate was 80% at the highest resolution (300×300 pixels). The recognition rate gradually increases as the resolution of MHI decreases and the beset recognition rate is 96.7 % at the lowest MHI resolutions (25×25 pixels).

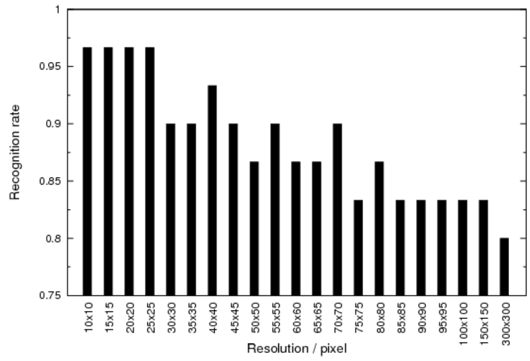

Figure 3. The recognition rates obtained by changing the resolutions of MHI.

We evaluated the recognition rates by changing duration time to calculate MHI. Since the MHI are modified by changing the duration time, it probably has influence on the recognition rates. The results are shown in Fig. 4. The resolutions of MHI were 10×10, 30×30 and 90×90 pixels, respectively. The method of recognition experiment was same as Section 3.1. From Fig. 4, the maximum recognition rates at each resolution were equal to the results of Fig. 3. The minimum recognition rate for 90×90 pixels was 83.3 %. The maximum recognition rate for 10×10 pixels was 96.7 % when the duration times were 100, 500, 1000 and 1500 ms. The maximum recognition rate for 30×30 pixels was 96.7 % when the duration times were 10, 100 and 1500 ms. The maximum recognition rate for 90×90 pixels was 96.7 % when the duration times was 100 ms. From Fig. 4, the maximum recognition rate was obtained when the duration time was 100 ms at each resolution. If the duration time is shorter or longer than 100 ms, the recognition rates become lower. These results suggest that too long duration time causes the decrease of the motion recognition rates.

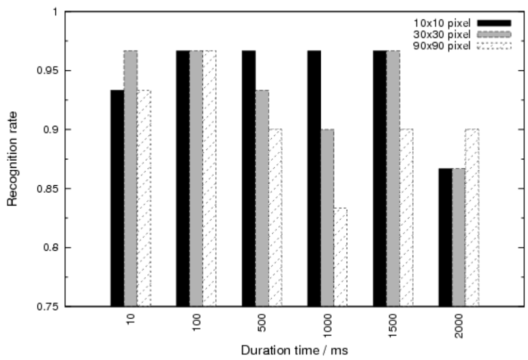

Figure 4. The recognition rates obtained by changing duration time.

4. Conclusion and Future works

The proposed method was able to recognize the pitchers 100 % correctly and it was also able to recognize whether there was a runner on a base with the recognition rate 96.7 %. These results show the effectiveness of the proposed LHAC-MHI features.

In the runner recognition, the better resolution rates were obtained when the resolutions of HMI were smaller than 25×25 pixels. The reason is that higher resolutions are probably affected by noises. This means that the necessary computational cost can be reduced by using low resolution MHI.

In general, the recognition rates can be improved by using the HLAC features extracted from multi-resolution images. How to combine the HLAC features extracted from multi-resolution MHI is for future work.

To investigate the influence of the duration time to calculate MHI, we have done experiments to measure the recognition rates by changing the duration time. From these experimental results, it is noticed that we have to adequately select the duration time to get better recognition rates. So we have to develop a method to select the optimal duration time automatically.

The purpose of this paper is to show the effectiveness of the proposed HLAC-MHI features, we used it and DP matching. These can be replaced to improve the recognition performance. For example, we can use linear discriminant analysis (LDA) or Hidden Markov Model (HMM).

References

[1] L.W. Campbell, D.A. Becker, A. Azarbayejani, A.F. Bobick, and A. Pentland, "Invariant features for 3-D gesture recognition," *Proc. International Conf. on Automatic Face and Gesture Recognition (FG96)*, pp.157-162,1996.

[2] K. Mase, "Recogniton of facial expression from optical flow," EICE transactions, vol.E74, no.10, pp.3474-3483, 1991.

[3] T. Kurita, Y. Kobayashi, and S. Mishima, "Higher order local autocorrelation features of PARCOR images for gesture recognition," *Proc. of International Conf. on Image Processing (ICIP97)*, pp.722-725, 1997.

[4] T. Kurita and S. Hayamizu, "Gesture recognition using HLAC features of PARCOR images and HMM based recognizer," *Proc. of International Conf. on Automatic Face and Gesture Recognition (FG98)*, pp.422-427, 1998.

[5] T. Kurita and S. Hayamizu, "Gesture recognition using HLAC features of PARCOR images," *IEICE Trans. on Inf. & Syst.*, Vol.E86-D, No.4, pp.719-726, 2003.

[6] N. Otsu and T. Kurita, "A new scheme for practical, flexible and intelligent vision systems", *Proc. IAPR Workshop on Computer Vision*, pp.431-435, 1988.

[7] N. Otsu, "An Approach to Adaptive and Trainable Vision", *IEIC Technical Report*, pp.171-172, 2006.

[8] T. Kobayashi and N. Otsu, "Action and Simultaneous Multiple-Person Identification Using Cubic Higher-order Local Auto-Correlation", *Proc.17 th. Int. Conf. on Pattern Recognition (ICPR'04)*, pp. 741-744, 2004.

[9] T. Nanri and N. Otsu, "Unsupervised abnormality detection in video surveillance," *Proc. IAPR Conf. on Machine Vision Applications (VMA2005)*, pp.574-577, 2005.

[10]A. Bobick and J. Davis, "Real-time recognition of activity using temporal templates," *Prof. of IEEE Workshop on Applications of Computer Vision (WACV '96)*, pp. 39-42, 1996.

[11]J. A.Mclaughlin and J. Raviv, "Nth-order autocorrelations in pattern recognition," *Inf. Control*, Vol.12, pp.121-142, 1968.