Human Action Recognition by Employing DWT and Texture

V. Thanikachalam and K.K. Thyagharajan

Abstract Human action recognition is a very challenging task due to the great variability with which different people may perform the same action. It involves in the development of applications such as automatic monitoring, surveillance, and intelligent human–computer interfaces. We propose an action recognition scheme to classify human actions based on positive portion using template-based approach from a video. We first define the accumulated motion image (AMI) using frame differences to represent the spatiotemporal features of occurring actions. Then, the direction of motion is found out by computing motion history image (MHI). Texture and spatial information are extracted from AMI and MHI using (LBP) local binary pattern and (DWT) discrete wavelet transform, respectively. The detection of object and extraction of moving objects are done by feature extraction over LBP and DWT. The feature vectors are computed by employing the seven Hu moments. The system is trained using nearest neighbor classifier, and the actions are classified and labeled accordingly. The experiments are conducted on Weizmann dataset.

Keywords Accumulated motion image · Discrete wavelet transform · Human action recognition · Local binary pattern · Motion history image · Nearest neighbor algorithm

1 Introduction

Due to the popularization of surveillance cameras and personal video devices, video-based human motion analysis and recognition have become a highly active area in computer vision. Human action recognition is difficult for many reasons,

V. Thanikachalam (✉)
SSN College of Engineering, Chennai, India
e-mail: thanikachalamv@ssn.edu.in

K.K. Thyagharajan
RMD Engineering College, Chennai, India
e-mail: kkthyagharajan@yahoo.com

such as high dimension of video data, intra-class variability caused by scale, viewpoint and illumination changes, low resolution, and video quality. Human action recognition is the process of identifying human actions that occur in the video sequences. Its application is involved in surveillance footage, user interfaces, robotics, automatic video organization, patient monitoring systems, athletic performance analysis, etc. Even though this topic has been studied for several years, it has some problem in classifying the actions performed by human accurately.

2 Review of Related Work

Various surveys have been done in this area on recognition methodologies and approach-based and general overview-based taxonomies [1, 2]. Tracking motion of human and action recognition from the sequence of images is analyzed [3]. The extension of [3] in the development of motion analysis has been surveyed by Wang et al. [4]. This paper is organized in hierarchical manner which describes issues caused during detection, tracking, and behavior understanding within human motion analysis system. Kim et al. [5] proposed the scheme over an ordinal measure of accumulated motion in a single query action video without considering whole silhouette of human body as in the earlier [6].

A set of activities is described using motion template system in [7]. Various methods which employ MHI have been proposed in [8] and provide variants and its applications of this method. The combined prompt of MHI and MEI is proposed in [9] for representation of motion based on optical flow-based directional history and energy image templates. To extract the judicial information from given video sequences, Shao et al. [10] proposed a system using wavelet transform-based descriptors for representation and recognition of human action. A multiscale representation has been proposed in [11] by comparing the efficiency with Gabor filter and discrete wavelet transform (DWT) which applied prior to dimensionality reduction. It results Gabor has the highest efficiency in recognition.

The local binary pattern (LBP) has been used from past 10 years, and it is widely used for texture classification proposed in [12–14].

3 Proposed Work

The proposed method consists of the following stages: accumulated motion image (AMI), motion history image (MHI), DWT, LBP, feature extraction, training, and classification. The overall procedure is shown in the flow diagram Fig. 1.

Fig. 1 Flow diagram
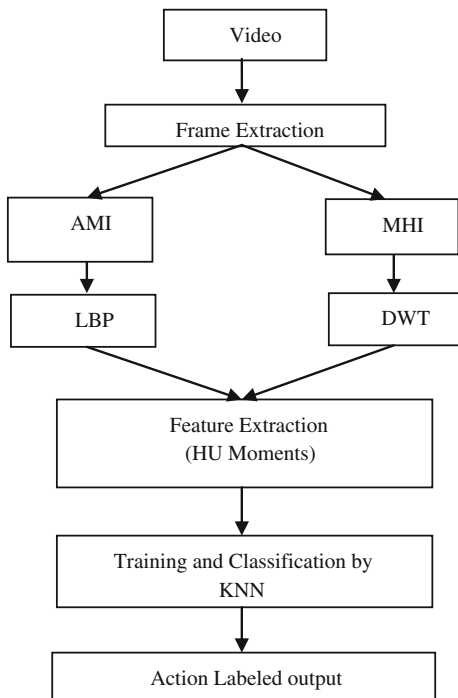
3.1 Accumulated Motion Image (AMI)

In the proposed system to represent the spatiotemporal features of human actions, we define AMI and it is computed by using frame differences. AMI is computed using frame differences as in Eq. 1

$$\text{AMI} = \frac{1}{T} \sum_{t=1}^{T} |D(x, y, t)|$$ (1)

where $D(x, y, t) = I(x, y, t) - I(x, y, t - 1)$ and T denotes the total number of frames present in a single action video.

Figure 2 shows the AMI computed for a jacking video and one-hand wave action.

3.2 Motion History Image (MHI)

MHI is extensively used in action recognition research areas. It provides motion shape information of a video to recognize actions.

**Fig. 2** AMI for jack and one-hand wave action

To represent how the motion occurs in a video, form a motion history image (MHI) is computed by using a simple replacement and decay operator as in Eq. 2

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

where τ is the current time stamp and D is the absolute value of silhouette difference between frames t and $t-1$. The result will be the scalar-valued image where brighter pixel shows the most recently occurred action.

Figure 3 shows the MHI computed for a jacking action and one-hand wave action video

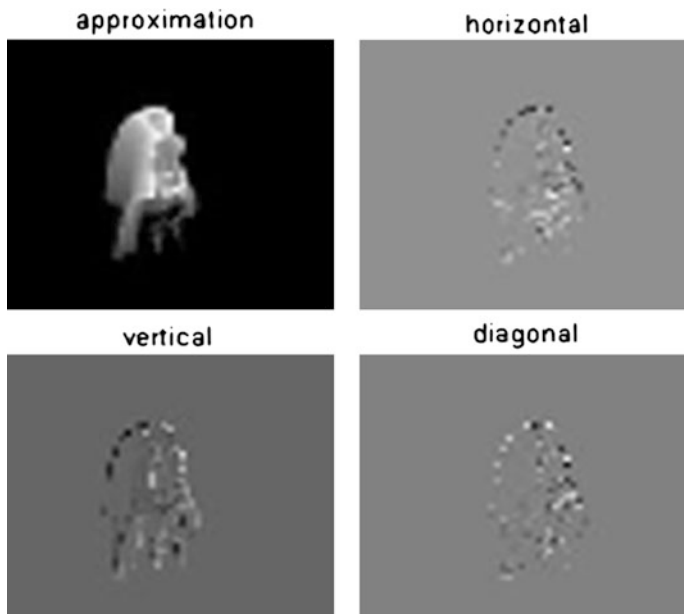**Fig. 3** MHI for jacking and one-hand wave action
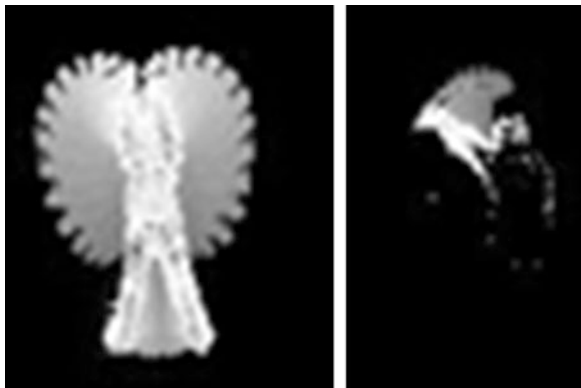
Fig. 4 DWT of MHI in different directions

3.3 Discrete Wavelet Transform (DWT)

DWT is used to extract the spatial features and for dimensionality reduction. In the proposed system, we use Haar wavelet filter for decomposition of the MHI image.
Figure 4 shows the output of the DWT.

3.4 Local Binary Pattern (LBP)

LBP is applied over AMI to extract the edge information. In texture analysis, the sampling is done on equally spaced sampling points P on a circular neighborhood with radius R. The value of the LBP code of the pixel is given in P-bit binary number as in Eq. 3

$$\text{LBP}_{P,R} = \sum_{P=0}^{P} S(g_p - g_c)2^p \quad s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where

g_c gray value of center pixel (x_c, y_c, t_c)
g_p gray value at p_{plane} sampling points

Figure 5 shows the output of the LBP.

Fig. 5 LBP of AMI

3.5 Feature Extraction

In the present work, the moment invariants are evaluated using central moments of the image function $I(x, y)$ up to third order. Regular moments are defined as in Eq. 4

$$M_{pq} = \sum_x \sum_y x^p y^q \qquad (4)$$

where for $p, q = 0, 1, 2, \ldots$ and is M_{pq} the order moment $(p + q)$th of the continuous image function $I(x, y)$. The central moments of order up to 3 are calculated using Eq. 4 which is invariant under translation.

3.6 Training and Classification

The system has to be trained with some classification algorithm which helps and finds the exact match of test actions. In the proposed system, K-nearest neighbor (KNN) classification method is applied. KNN algorithm is a part of supervised learning that has been used in many applications in the field of data mining, statistical pattern recognition, and many others.

KNN is a method for classifying objects based on closest training examples in the feature vector. An object is classified by a majority vote of its neighbors. K is always a positive integer. If $K = 1$, it is then called as nearest neighbor (NN). The neighbors are taken from a set of objects for which the correct classification is known.
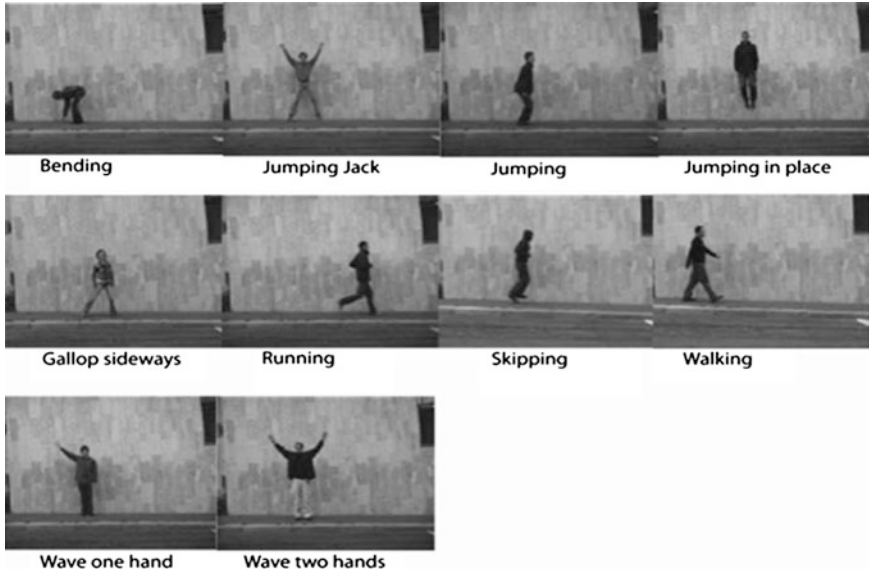
Fig. 6 Some sample of 10 action classes in Weizmann dataset

3.7 Weizmann Dataset

The Weizmann dataset has been used in the proposed system which consists of large number of subjects and actions. It includes 81 low-resolution videos from nine different people, each performing 10 natural actions. A sample is shown in Fig. 6.

4 Results and Discussion

The proposed system is worked out with Weizmann dataset consists of 8 actions of different persons. The actions including bend, jumping jack, jumping, walk, run, skip, gallop side, and wave.

Number of actions taken for testing = N

The total classification rate of the proposed system is calculated as follows

$$C = \frac{(N - \text{Actions wrongly recognized})}{N}$$

The percentage of video giving correct output is 90 %, and the percentage of video giving wrong output is 10 %. The accuracy given above is obtained by using

both accumulated motion image and MHI with LBP and DWT, respectively. It gives better result when both the techniques are concatenated rather than applied separately. This proposed method is able to recognize 9 out of 10 actions.

5 Conclusions

AMI evaluation helps us in handling the variations of appearance and clearly shows the moving pixels. MHI is defined, and its determination gives direction of motion. LBP gives shape and texture information of AMI and also helped to improve recognition accuracy. DWT aids in extracting spatial features of MHI. Then, the feature vectors of LBP and DWT obtained by Hu seven moments and distance metric calculated using KNN classifier and the actions are labeled accordingly. The proposed system gives the high performance when compared with previous proposals which uses KNN. The computation is simpler and less time consuming.

References

1. J.K. Aggarwal, M.S. Ryoo, Human activity analysis: a review. ACM Comput. Surv. **43**(3), 16 (2011)
2. T. Moeslund, E. Granum, A survey of computer vision based human motion capture. Comput. Vision Image Understand. **81**, 231–268 (2001)
3. J.K. Aggarwal, Q. Cai, Human motion analysis: a review. Comput. Vision Image Understand. **7**(3), 428–440 (1999)
4. L. Wang, W. Hu, T. Tan, Recent developments in human motion analysis. Pattern Recognit. **36**, 585–601 (2003)
5. W. Kim, J. Lee, M. Kim, D. Oh, C. Kim, Human action recognition using ordinal measure of accumulated motion. EURASIP J. Adv. Signal Process. **2010**, 1–12 (2010)
6. V.H. Chandrashekhar, K.S. Venkatesh, Action energy images for reliable human action recognition, in *Proceedings of the Asian Symposium on Information Display* (*ASID '06*), pp 484–487 (2006)
7. J. Davis, A. Bobick, in *The representation and recognition of action using temporal templates*. IEEE Conference on Computer Vision and Pattern Recognition (1997)
8. M.A.R. Ahad, J.K. Tan, H. Kim, S. Ishikawa, in *Human Activity Analysis: Concentrating on Motion History Image and Its Variants*. ICROS-SICE International Joint Conference (2009)
9. M.A.R. Ahad, J.K. Tan, H. Kim, S. Ishikawa, Action recognition by employing combined directional motion history and energy images. Image Vision Comput. 73–78 (2010)
10. L. Shao, R. Gao, Y. Liu, H. Zhang, Transform based spatiotemporal descriptors for human action recognition. Neurocomputing **74**, 962–973 (2010)
11. R. Shrivastava, A. Pratap Singh, Analysis and performance of three methods of human action recognition. Int. J. Adv. Res. Electr. Commun. Eng. (IJARECE). **19**(3), 72 (2012)
12. T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
13. T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–984 (2002)
14. P. Guo, Z. Miao, Motion description with local binary pattern and motion history image: application to human motion recognition, in *IEEE International Workshop on Haptic Audio Visual Environments and their Applications,* pp 18–19(2008)