Research Journal of Applied Sciences, Engineering and Technology 11(10): 1159-1163, 2015
DOI: 10.19026/rjaset.11.2131
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: July 19, 2015 Accepted: August 18, 2015 Published: December 05, 2015

Research Article

Human Action Recognition Using Temporal Partitioning of activities and Maximum Average Correlation Height Filter

[1]V. Thanikachalam and [2]K.K.Thyagharajan
[1]SSN College of Engineering, Chennai, Tamil Nadu-603110,
[2]RMD Engineering College, Chennai, India

Abstract: We proposed a method for Human action Recognition. It is based on the construction of a set of templates for each activity. Each template is constructed based on the Accumulated Motion Image of the Video. Each template contains where motion has occurred in the video. FFT Transform is applied to each template. A 3D Spatiotemporal Volume is generated for each class. A Single action Maximum average Correlation height Filter is generated for each class. The filter is applied to the test video and using the threshold the actions are classified. The experiments are conducted on Weizmann dataset.

Keywords: Accumulated motion image, Fourier transform, human action recognition, maximum average correlation height filter

INTRODUCTION

Human motion analysis and recognition have become a highly active area in computer vision, due to the presence of Surveillance cameras and personal video devices. Yet effective solution is not obtained because of 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. The application domains are in Surveillance footage, User-interfaces, Robotics, Automatic video organization, patient monitoring systems, athletic performance analysis etc. Classification of human actions is not done accurately because of several reasons.

Aggarwal and Ryoo (2011) provided an overview of the current approaches to Human activity Recognition. They have explored the various methodologies that is used in human action recognition. Kim *et al*. (2010) used AMI and Energy Histograms for Human Action Recognition. Davis and Bobick (1997) computed hu moments of Motion Energy Image and Motion History Images to create action template. Ahad *et al*. (2009, 2010) presented all important variants of the Motion History image Method and suggests some areas for further research. Chandrashekhar and Venkatesh (2006) construct the Eigen Activity Space by performing PCA on AEIs of various activities and use it for recognition. Shrivastava and Singh (2012) analysed the performance of three methods of human action recognition. Ping and Zhenijiang (2008) adopt the ideas of spatiotemporal analysis and the use of local features for motion description and they are computationally simple and suitable for various application. Mahalanobis *et al*. (1987) created a new category of spatial filters MACE that produces sharp output correlation peaks with controlled peak values. Rodriguez *et al*. (2008) introduced MACH filter that generate a single action template by using all the frames in the training and testing Videos. The objective of this research work is to propose a Human Action Recognition method Using Temporal Partitioning of activities and Maximum average Correlation Height Filter for Weizmann dataset.

MATERIALS AND METHODS

Recently a lot of attention is shown towards analyzing human actions in spatiotemporal space instead of analyzing each frame. The proposed method is computationally simple.

Materials used in this research work: The experiments were conducted with Weizmann dataset. The proposed method was implemented using Matlab.

Proposed method: Our proposed methodology for human action recognition is shown in Fig. 1. It has the main components of the proposed recognition. As shown in the block diagram for the Training Video 4 Temporal segments are constructed and 3D FFT is applied and MACH filter is constructed. In testing Videos 2 temporal segments are used out of 4 temporal segments and 3D FFT is applied and MACH filter is applied and the peak value is calculated. The peak value

Corresponding Author: V. Thanikachalam, SSN College of Engineering, Chennai, Tamil Nadu-603110, India

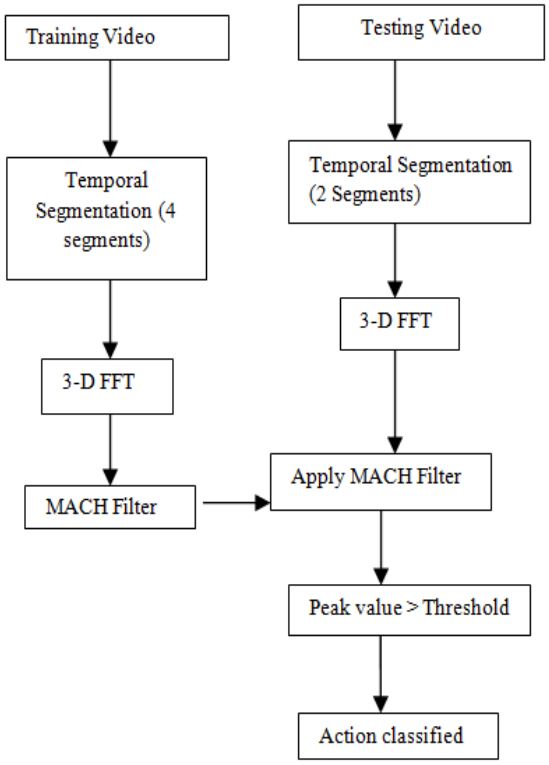

Fig. 1: Main structure of the proposed system

is compared with the threshold and actions are classified.

Temporal segmentation of activities: An activity can be performed by the same person or by different person in different ways because of the variation in the speed of the action. Even the same person can vary the speed during the activity is performed. So this temporal variability in performing the activity initiates the deployment of methods which are robust to variations. Because of this reason we divided the given action video in to four stages and for each stage a template is constructed.

Accumulated Motion Image (AMI): In the proposed system to represent the spatio-temporal 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):

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

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. AMI gives where motion has occurred in the video. It captures the pose details in the given activity.

Algorithm for temporal segmentation:

- An activity video is divided in to four equal temporal segments.

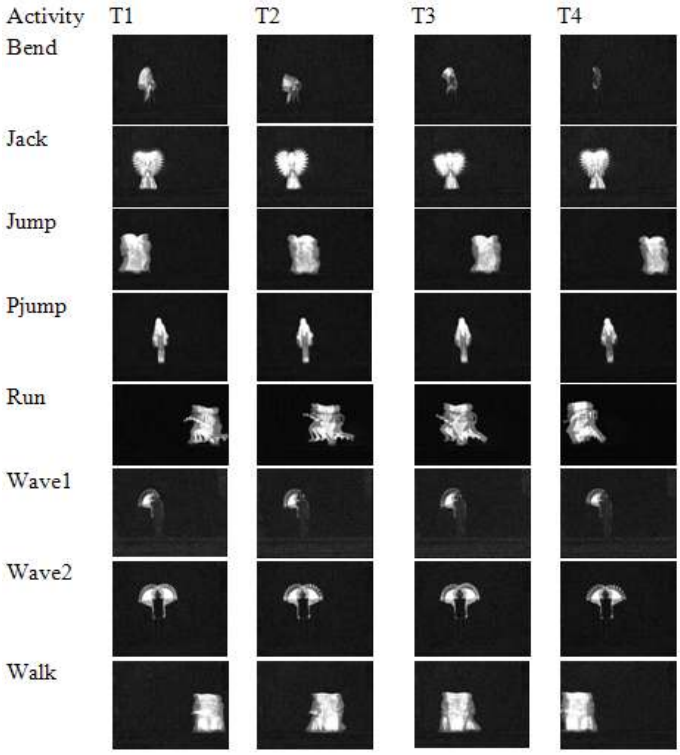

Fig. 2: Temporal segmentation for each action

- AMI is calculated for each temporal segment where each segment has equal number of frames.
- Each AMI act as a template. So four templates has been generated for each activity.
- Template 2 and 3 provides more information when compared to 1 and 4.
- The four stage template will be referred to as spatiotemporal profiles.
- Temporal segmentation is shown in Fig. 2.

Fast Fourier Transform (FFT): The FFT is the sampled Fourier Transform and therefore does not contain all frequencies forming an image, but only a set of samples which is large enough to fully describe the spatial domain image. The number of frequencies corresponds to the number of pixels in the spatial domain image.

Maximum average correlation height filter: The MACH filter is used in Object Classification, Palm print identification problems. The MACH filter produces a single composite template from the instances of a class by optimizing the four metrics:

- The Average Correlation Height (ACH)
- The Average Correlation Energy(ACE)
- The Average Similarity Measure(ASM)
- The Output Noise Variance (ONV)

The process results in a 2D template that gives the shape or appearance of an object in the video.

MACH filter for the action: The Process of training the MACH filter with the creation of Spatio-temporal volumes by concatenating the templates of an action. A 3D FFT operation is performed to represent each spatio-temporal volumes in the frequency domain as shown in Eq. (2):

$$F(u, v, w) = \sum_{t=0}^{N-1}\sum_{y=0}^{M-1}\sum_{x=0}^{L-1} f(x, y, t)\exp\left(-j2\Pi\left(\frac{uv}{L}+\frac{vy}{M}+\frac{wt}{N}\right)\right) \quad (2)$$

where, f(x, y, t) is the volume corresponding to the templates of the input sequences.
F(u, v, w) is the frequency volume.

L = Number of columns
M = Number of rows
N = Number of Frames

The Resultant 3D FFT matrix is converted in to a Single column vector denoted by x_i.

The MACH filter is created in the frequency domain as follows in Eq. (3):

$$h = (\alpha C+\beta D_x+\chi Sx)^{-1} m_x \quad (3)$$

where,
m_x = The mean of all x_i
h = The filter in the frequency domain
C = The diagonal noise covariance matrix
α = The Standard deviation parameter

D_x represents average power spectral density of the training video and is defined in Eq. (4):

$$D_x = \frac{1}{N_e}\sum_{i=1}^{N_e} X_i^* X_i \quad (4)$$

where,
x_i = A diagonal matrix
* = The conjugate operations

S_x is the diagonal average similarity matrix defined as in Eq. (5):

$$S_x = \frac{1}{N_e}\sum_{i=1}^{N_e}(x_i - M_x)^*(x_i - M_x) \quad (5)$$

M_x is a diagonal matrix. α, β, χ are the parameters that can be set to obtain the performances. Finally the 1-D filter h is designed.

Action classification: The MACH filter is applied to the test video in which the 2nd and 3rd template alone is used. Here the entire test video is not used with the MACH filter. The response is calculated as shown in Eq. (6):

$$C(l, m, n) = \sum_{t=0}^{N-1}\sum_{y=0}^{M-1}\sum_{x=0}^{L-1} s(l+x, m+y, n+t)H(x, y, t) \quad (6)$$

where, S is the spatio temporal volume of the test video. H is the MACH filter. The response C is normalized and its value lies within 0 and 1. The peak value of the response filter is compared with the threshold. If the response of the filter is greater than the threshold we inferred the action has occurred and it is classified.

Weizmann dataset: The Weizmann dataset has been used in the proposed system which consists of relatively larger in terms of the number of subjects and actions. It includes 81 low-resolution videos from 9 different people, each performing 10 natural actions. A sample is shown in the Fig. 3.

RESULTS AND DISCUSSION

The proposed system is worked out with Weizmann dataset consists of 10 actions of different persons. The actions including bend, jumping jack, jumping, walk, run, skip, gallop side and wave.

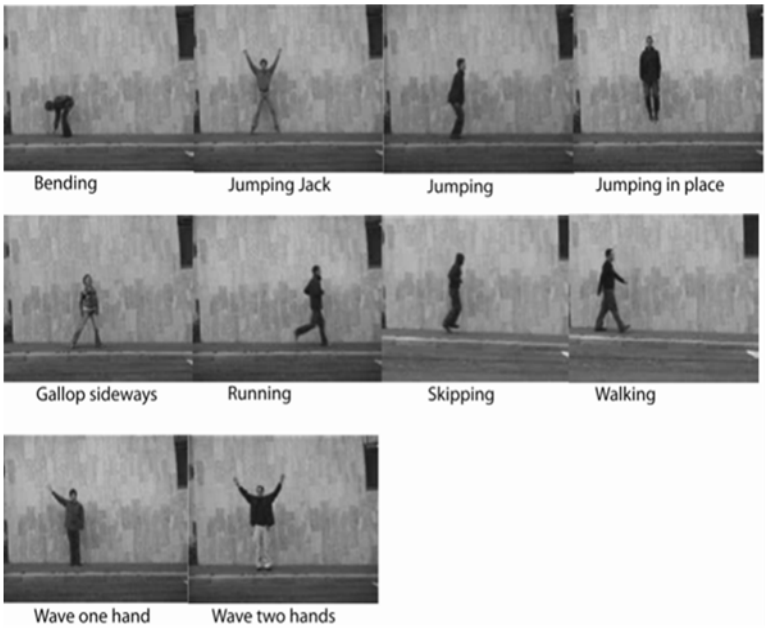

Fig. 3: Some sample of 10 action classes in Weizmann dataset

Table 1: Activity and threshold matrix

Activity	Threshold
Bend	0.84
Jack	0.87
Jump	0.56
Pjump	0.80
Run	0.66
Wave1	0.85
Wave2	0.86
Walk	0.64
Side	0.60
Skip	0.40

The Threshold that is used to classify the action is shown in the Table 1 for Weizmann dataset:

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 92% and the percentage of video giving wrong output is 8%.The accuracy given above is obtained by using Accumulated motion image with MACH Filter. This proposed method is able to recognize 9 out of 10 actions.

CONCLUSION

An activity is divided in to four temporal segments. AMI is generated for each temporal segment and it acts a template for each segment. The templates along with MACH filter is used for Human action Recognition. The computation is simpler and less time consuming as no classifier is used in the system. 3-D FFT is applied only to the templates and not for the entire video. The system performance can be enhanced by fusing multiple features.

REFERENCES

Aggarwal, J.K. and M.S. Ryoo, 2011. Human activity analysis: A review. ACM Comput. Surv., 43(3).

Ahad, M.A.R., J.K. Tan, H. Kim and S. Ishikawa, 2009. Human activity analysis: Concentrating on motion history image and its variants. Proceeding of the ICROS-SICE 2009, pp: 5401-5406.

Ahad, M.A.R., J. Tan, H. Kim and S. Ishikawa, 2010. Action recognition by employing combined directional motion history and energy images. Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW, 2010), pp: 73-78.

Chandrashekhar, V.H. and K.S. Venkatesh, 2006. Action energy images for reliable human action recognition. Proceeding of the Asian Symposium on Information Display (ASID '06), pp: 484-487.

Davis, J. and A. Bobick, 1997. The representation and recognition of action using temporal templates. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, pp: 928-934.

Kim, W., J. Lee, M. Kim, D. Oh and C. Kim, 2010. Human action recognition using ordinal measure of accumulated motion. EURASIP J. Adv. Sig. Pr., Vol. 2010, Article ID 219190.

Mahalanobis, A., B.V.K. Vijaya Kumar and D. Casasent, 1987. Minimum average correlation energy filters. Appl. Opt., 26(17): 3633-3640.

Ping, G. and M. Zhenijiang, 2008. Motion description with local binary pattern and motion history image: Application to human motion recognition. Proceeding of the IEEE International Workshop on Haptic Audio Visual Environments and Games (HAVE, 2008), pp: 171-174.

Rodriguez, M.D., J. Ahmed and M. Shah, 2008. Action mach a spatio-temporal maximum average correlation height filter for action recognition. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR, 2008), pp: 1-8.

Shrivastava, R. and A.P. Singh, 2012. Analysis and performance of three methods of human action recognition. Int. J. Adv. Res. Electron. Commun. Eng. (IJARECE), 1(3).