ISSN: 2321-7782 (Online)
Volume 3, Issue 11, November 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Automatic Understanding of Human Actions in Videos

Thanikachalam V[1]
Information Technology
SSN College of Engineering
Chennai, India

Thyagharajan K K[2]
Electronics and Communication Engineering
RMD Engineering College
Chennai, India

Abstract: ***Human action recognition has been the centre of interest of many researchers over the last two decades. Specifically making a computer that recognizes and understands human actions is the most intriguing research in computer vision and artificial intelligence. This paper analyzes the various methods of recognizing actions with single layer space-time approach. These methods help to create a framework for speeded up human-computer interactions.***

Keywords: ***Accumulated Motion Image, Motion History Image, Local Binary Pattern, Discrete Wavelet Transform, FFT, KNN, Correlation Filter***

I. INTRODUCTION

The purpose of Human action recognition is to automatically understand the activities performed by a human in a video. The Human activity is categorized in to four levels as gestures, actions, interactions and group activities. Gestures are elementary movements of a person's body part like raising the hand is an example of gestures. In [1] a frame work has been developed to identify the sports events from an image (2D) using Scale Invariant Transform and Bag of Visual Words. Actions are single person activities such as jumping, single hand waving and running. Interactions are that has more than one person or object. Persons fighting or carrying a suitcase are examples of Interactions. The activities performed by conceptual groups composed of multiple persons or objects are group activities. Group meeting and group fighting are examples of group activities. Human activity recognition methodologies [2] [3] are classified in to two categories as Single-layered approaches [4] and Hierarchical approaches. Single-layered approaches are purely based on sequences of images. Hierarchical approaches represent high-level human activities by describing them as sub-events. Single layer approach is divided in to two classes as Space time approaches and sequential approaches. In Space time approaches the similarity is measured between two volumes. Each action is represented with a template composed of two dimensional images. The two dimensional images are Accumulated Motion Image (AMI) and a Scalar-valued Motion-History Image (MHI). The two images are constructed from a sequence of foreground images which produces 2-D projections of the original 3-D Space-time volumes. The paper is organized as follows Section 2 gives detailed explanation about the three methods implemented, Section 3 details the Dataset, section 4 analyses the results, section 5 is conclusion. The system is tested on the existing Weizmann dataset.

II. METHODS FOR HUMAN ACTION RECOGNITION

Action recognition schemes were developed to classify human actions based on positive portion using template based approach from a Video. We first define the accumulated motion image (AMI) [5] 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). AMI is computed by using frame differences as in Equation 1

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

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. The x and y represents the position of the pixel. The Intensity value of the pixel is used to calculate the AMI.

Fig. 1 AMI for Two Hand Wave Actions

Motion History Image (MHI) [6] is extensively used in action recognition research areas [7]. It provides motion shape information of a video and it is computed by using a simple replacement and decay operator as in Equation 2

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

Where τ is the current timestamp 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.

Fig. 2 MHI for Bend Actions

A. Employing LBP and DWT

Local Binary Pattern (LBP) is applied over AMI to extract the edge information. The meaningful information is present in the boundary of the AMI. The basic idea is to summarize the local structure in an image by comparing each pixel with its neighborhood. Take a pixel as center and threshold its neighbors against. If the intensity of the center pixel is greater-equal its neighbor, then denote it with 1 and 0 if not. Local Binary Pattern (LBP) is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. Texture and spatial information are extracted from AMI and MHI using Local Binary Pattern (LBP) [8] and (Discrete Wavelet Transform) DWT respectively.

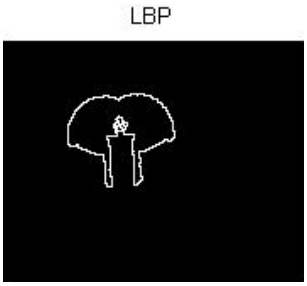

Fig. 3 LBP Based Texture Identification Using AMI for Two Hand Wave Action

DWT [9] is used to extract the spatial features and for dimensionality reduction. In the proposed system the Haar wavelet filter is used for decomposition of the MHI image. The Haar wavelet transform decomposes a signal into a time-frequency field based on the Haar wavelet function basis. For discrete digital signals, the discrete wavelet transform can be implemented efficiently by Mallat's fast algorithm. We keep the low frequency part of the Haar wavelet transform, so that total dimension of the combined feature vector is lower than that of the original MHI feature. From the final template images, the corresponding feature vectors are computed by employing the seven Hu moments. Now we get 7 features for each DWT and LBP, and then

concatenate it in to 14 features. In the proposed system K-Nearest Neighbor (KNN) classification method is applied for training and classification.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.

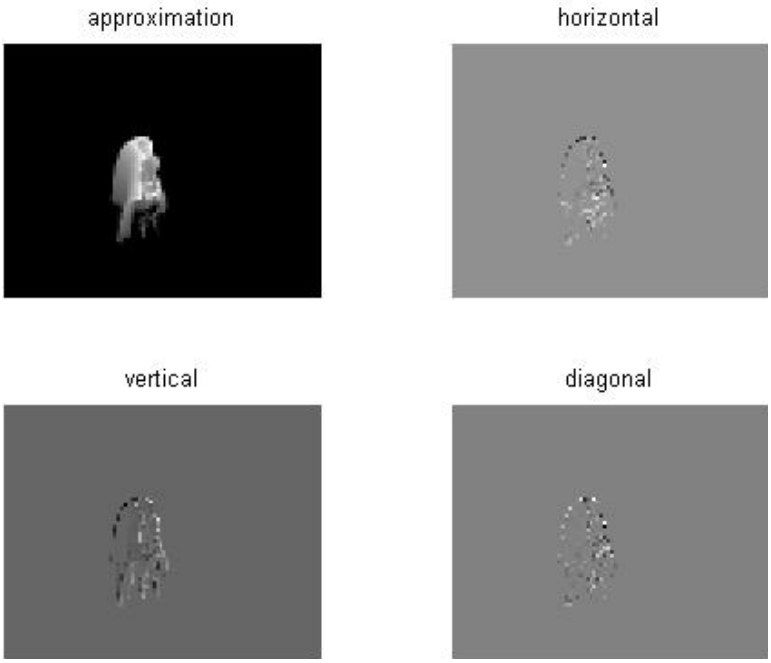

Fig 4. DWT of MHI for Bend action

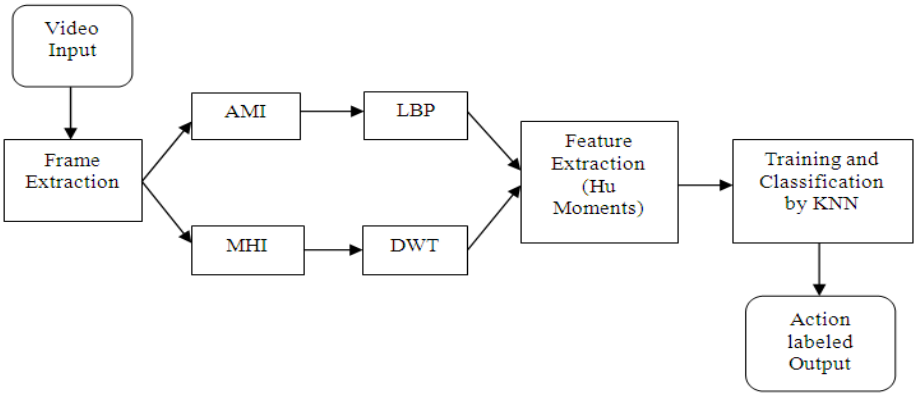

Fig. 5 Block Diagram of the system employing LBP and DWT

B. Human Action Recognition Using MHI and Correlation Filter

In this method the FFT is applied to the MHI. Minimum average correlation energy (MACE) filter [10] minimizes the average correlation energy (ACE) of the correlation outputs due to the training images simultaneously satisfying the correlation peak constraints at the origin. MACE filters are obtained by forcing the average cross-correlation plane energy to minimum for the training images with hard constraints at the origin of the plane to yield specific value. It implies that resulting filter gives cross-correlation plane. However, in practice we do not get the exact delta function but resulting peaks are very sharp and provide a good measure for discrimination between authentic and impostor images. The MACE filter h in frequency domain is Equation 3

$$h = D^{-1} X \left(X^{+} D^{-1} X \right)^{-1} u \qquad (3)$$

Assume we have N training image each having P number of pixels. Then X is P×N matrix of the Fourier transform of the training images arranged in a lexicographic manner. D is P×P diagonal matrix of the average power spectrum of training images. And u is the pre-specified value at the origin of the correlation plane.

The correlation output should be sharply peaked and it should not exhibit such strong peaks for impostors. Calculate the Peak to Sidelobe Ratio(PSR) as shown in equation 4.

$$PSR = \frac{(Peak - Mean)}{Variance} \quad (4)$$

Match declared when PSR is large as training images are arranged in a lexicographic manner.

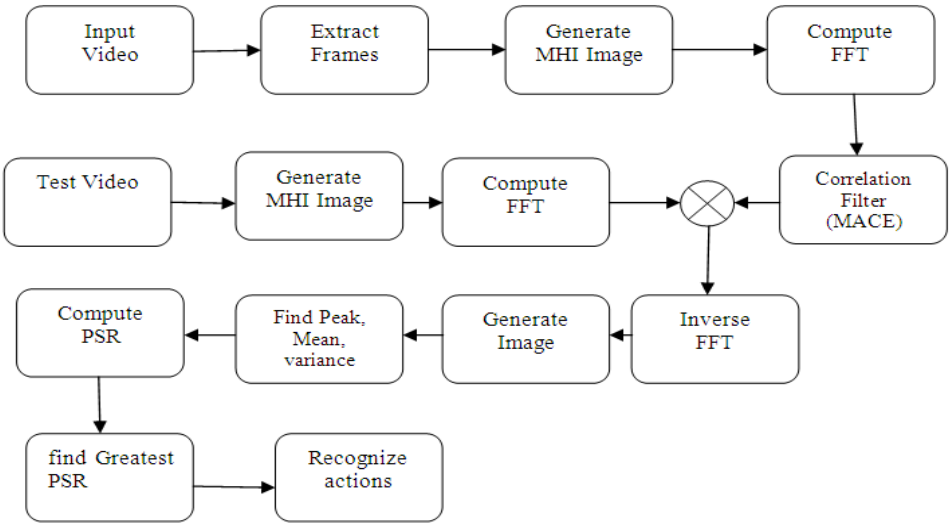

Fig.6 Block Diagram of the system using MHI and Correlation

C. Temporal partitioning of activities and Maximum average Correlation Height Filter.

It is based on the construction of a set of templates for each activity. The methodology is as follows

» Temporal segmentation is done for each 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 (MACH) is generated for each class.

» The filter is applied to the test video and using the threshold the actions are classified.

The segmentation is done as follows

1. An activity video is divided in to four equal temporal segments.
2. AMI is calculated for each temporal segment where each segment has equal number of frames.
3. Each AMI act as a template. So four templates has been generated for each activity.
4. Template 2 and 3 provides more information when compared to 1 and 4.
5. The four stage template will be referred to as spatiotemporal profiles.
6. Temporal segmentation is shown in Fig. 7.

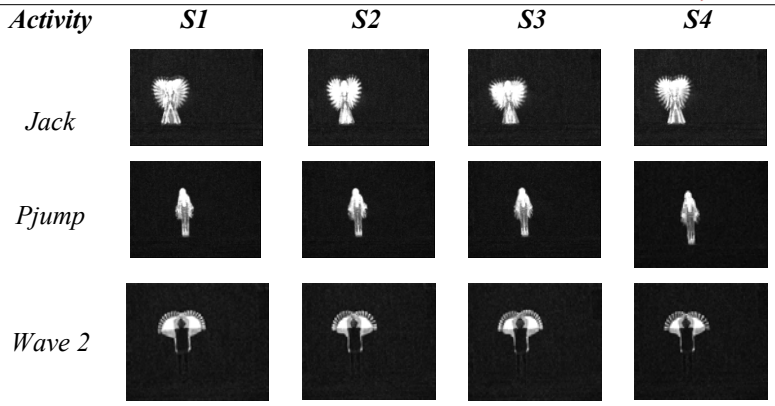

Fig.7 Temporal Segmentation for sample action

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

Table I Activity and Threshold matrix

Activity	Threshold
Bend	0.84
Jack	0.87
Pjump	0.80
Run	0.66
Wave2	0.86
Walk	0.64
Skip	0.40

III. WEIZMANN DATASET

The Weizmann human action dataset [11] [12] contains 93 sequences of nine actors performing ten different actions. Each sequence contains about 40 – 120 frames. It was recorded in the year 2005 for the purpose of studying new algorithms. The background is simple and a single person is performing the action in each frame. The actions performed by the humans are walking, running, jumping, galloping sideways, bending, one-hand waving, two hands waving, jumping in place, jumping jack, and skipping. The background in the video is static and the view point is static.

IV. EXPERIMENTAL RESULTS

The total classification rate of the proposed method can be defined as follows

$$C = \frac{(N - \text{No of misclassification})}{N} \times 100$$

Where N denotes the total number of videos used for comparison. The results are shown in Table 2

Table II Comparison of Classification Rates with Different Methods

Methods	Classification Rate	Description
AMI, MHI, LBP, DWT	90%	Number of pixels reduced, Low frequency pixel is considered. KNN Classifier is used
MHI, MACE	91%	Correlation filter is used. No classifier is needed. Classification is based on highest PSR value. Training data should be in Lexicographic order.
Temporal Segmentation, 3DMACH	92%	Each Class has one 3Dfilter irrespective of any number of actors. Classification is based on Threshold. No classifier is needed. Few segments of a video is used for testing.

V. CONCLUSION

The methods use different features which are simple and different classification algorithms are used which yields good results similar to the existing methodologies. All the methods that are discussed use the space-time volume approach which falls under the category of single-layered approaches that was used to model human activities. These methods can be improved by using SVM classifiers [13] and finally can be used for semantic video retrieval [14]. The 3D MACH filter is developed for each action irrespective of any number of actors performing the same action. The action videos are divided in to four segments and only two segments of the videos are used for testing purpose. No classifiers are used in this method and it gives better result than the other approaches. The limitations of this research work are space-time trajectories and space-time local features are not used in these approaches. They are suitable for recognizing periodic actions.

References

1. Thyagharajan K K, Nagarajan G, "Semantically Effective Visual Concept Illustration for Images", International Journal of Future Computer and Communication, Vol. 3, No. 2, pp. 124-128, 2014
2. J. K. Aggarwall and M. S. Ryoo, "Human Activity Analysis: A Review", ACM computing surveys, 2011
3. Yu-Gang Jiang, Subhabrata Bhattacharya, Shih-Fu Chang and Mubarak Shah, "High-level event recognition in unconstrained Video", Springer International Journal Multimedia Info Retr. 2012
4. Laptev.I, Marszalek.M, Schmid C and Rozenfeld.B, "Learning realistic human actions from Movies", IEEE Conference on computer Vision and pattern Recognition (CVPR), 2008.
5. Wonjun Kim, Jaeho Lee,Minjin Kim, Daeyoung Oh, and Changick Kim, "Human Action Recognition Using Ordinal Measure of Accumulated Motion", EURASIP Journal on Advances in Signal Processing Volume 2010.
6. J. Davis and A. Bobick, "The representation and recognition of action using temporal templates", Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1997
7. Md. Atiqur Rahman Ahad, J.K. Tan, H. Kim and S. Ishikawa, "Motion History Image: Its Variants and applications", Springer Machine Vision and Applications 23:255-281, 2012.
8. Vili Kellokumpu, Guoying Zhao and Matti Pietikainen, "Recognition of human actions using texture descriptors", Springer Machine Vision and Applications, 2007.
9. Hongying meng, Nick pears and Chris bailey "Motion feature combination for Human action Recognition in Video", Springer-Verlag Berlin Heidelberg, 2008.
10. A. Mahalanobis, B. V. K. Vijaya Kumar, andD. Casasent, "Minimum average correlation energy filters," Appl. Opt., vol. 26, no. 17, pp. 3633–3640, 1987.
11. Jose M.Chaquet, Enrique J. carmona, Antonio Fernansez-caballero, "A survey of video datasets for human action and activity recognition", Computer Vision and Image Understanding 117 (2013) 633-659, Elsevier, 2013.
12. Weizmanndataset:http://www.wisdom.weizmann.ac.il/~vision/ spaceTimeActions.html
13. Minu R I, Thyagharajan K K, "Automatic Image Classification Using SVM Classifier", CiiT International Journal of Data Mining and Knowledge Engineering, Vol. 3, No. 9, pp. 559 – 564, 2011.
14. Nagarajan G, Thyagharajan K K, "A Machine Learning Technique for Semantic Search Engine", Procedia Engineering, Elsevier ONLINE Journal, Vol. 38, pp. 2164-2171, 2012.

AUTHOR(S) PROFILE

V.Thanikachalam, is an Assistant Professor in the Department of Information Technology in SSN College of Engineering. He has 16 years of teaching experience. He received his B.E (Computer Science and Engineering) from Bharathidasan University, M.E.(CSE) from Anna University Chennai. He is currently pursuing Ph.D. (part time) at Anna University Chennai. He has published 4 papers in International journals and conferences. He has involved in many UG projects and PG Projects in the area of Image Processing.

K.K. Thyagharajan received his B.Eng. degree in Electrical and Electronics Engineering from PSG College of Technology, Madras University, India and received his M.Eng. degree in Applied Electronics from Coimbatore Institute of Technology, India in 1988. He also possesses a Post Graduate Diploma in Computer Applications from Bharathiar University, India. He obtained his Ph.D. degree in Information and Communication Engineering from College of Engineering Guindy, Anna University, India. He is in teaching profession for around three decades and served at various levels including Principal, Professor and Dean at various Engineering Colleges in Tamil Nadu-INDIA. He has written 5 books in Computing including "Flash MX 2004" published by McGraw Hill (India), which has served recommended as text and reference book by universities. He is a grant recipient of Tamil Nadu State Council for Science and Technology. He has been invited as chairperson and delivered special lectures in many National and International conferences and workshops. He is reviewer and editorial board member for many International Journals and Conferences. He is a recognized supervisor for Ph.D candidates and Master students at Anna University. He has published more than 90 papers in National & International Journals and Conferences. Three candidates have completed PhD under his supervision. His research interests include Computer Vision, Semantic Web, Image & Video Processing, Multimedia Streaming and e-learning. He is a life member of ISTE, CSI INDIA and also senior member and invited member in many professional associations. He has been recognized by Marquis Who's Who in the World for his contribution to the technical society and his biography has been published in its 25th Anniversary Edition.