Volume 3, Issue 7, July 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Recognition of Human Actions Using Motion History Information Extracted from the Compressed

**Renu Thakur**
*M.Tech Student (CSE)*
*RBIEBT,Sahauran, Punjab, India*

**Neha Mehan**
*Assistant Professor (CSE)*
*RBIEBT, Sahauran, Punjab , India*

**Namitakakkar**
*Assistant Professor (CSE)*
*RBIEBT, Sahauran, Punjab, India*

***Abstract— This paper describes a system for recognition of various human actions from compressed video based on motion history information. The encoded motion information available in the compressed MPEG stream is used to construct the coarse motion history image (MHI) and the corresponding MFH. The features extracted from the static MHI and MFH compactly characterize the spatio-temporal and motion vector information of the action. The features are extracted from the partially decoded sparse motion data, the computational load is minimized to a great extent. The extracted features are used to train K-Nearest Neighbour, Neural Network, Support Vector Machine, Bayes and Probalistic Neural Network for recognizing various set of human actions. The performance of each feature set with respect to various classifiers are analysed.***

***Keywords— Action Recognition; Feature Extraction; Classifiers; Feature Performance Analysis; Conclusion***

## I. INTRODUCTION

In the recent past, a technique for recognizing human actions from compressed video using (HMM) Hidden Markov Model, where the time-series features used for training the HMM are directly extracted from the motion vectors corresponding to each frame of the video. Though this approach has proven its ability to classify the video sequences, the extracted time series features are not suitable for other efficient classifiers such as K – Nearest Neighbors (KNN), Neural Network, Support Vector Machine, BayesClassifiers. In this paper we propose a technique for building coarse Motion History Information (MHI) and Motion Flow History (MFH) from the compressed video and extract features from these static motion history information for characterizing human action. The MHI gives the temporal information of the motion at the image plane and MFH quantifies the motion at the image plane. The features are extracted from MHI and MFH are used to train KNN, Bayes, Neural Network, SVM and Probalistic Neural Network classifiers for recognizing a set of various human actions. The various human actions are as follows:


JUMP


JUMP1

 


RUN


WAVE 1


WAVE 2


BEND

The encoded motion information available in the MPEG video is exploited for constructing the coarse MHI and MFH. These MHI and MFH represents the human action in a very compact manner. The motion information extracted from each frame of the compressed video is very sparse, they are sufficient to construct the MHI and MFH for representing the actions.

This work is motivated by a technique proposed by Davis and Bobick [11] where a view-Based approach is used to recognize actions. They have presented a method for recognition of temporal templates.

 

## II. SYSTEM OVERVIEW

The overview of the proposed system is shown in figure 1

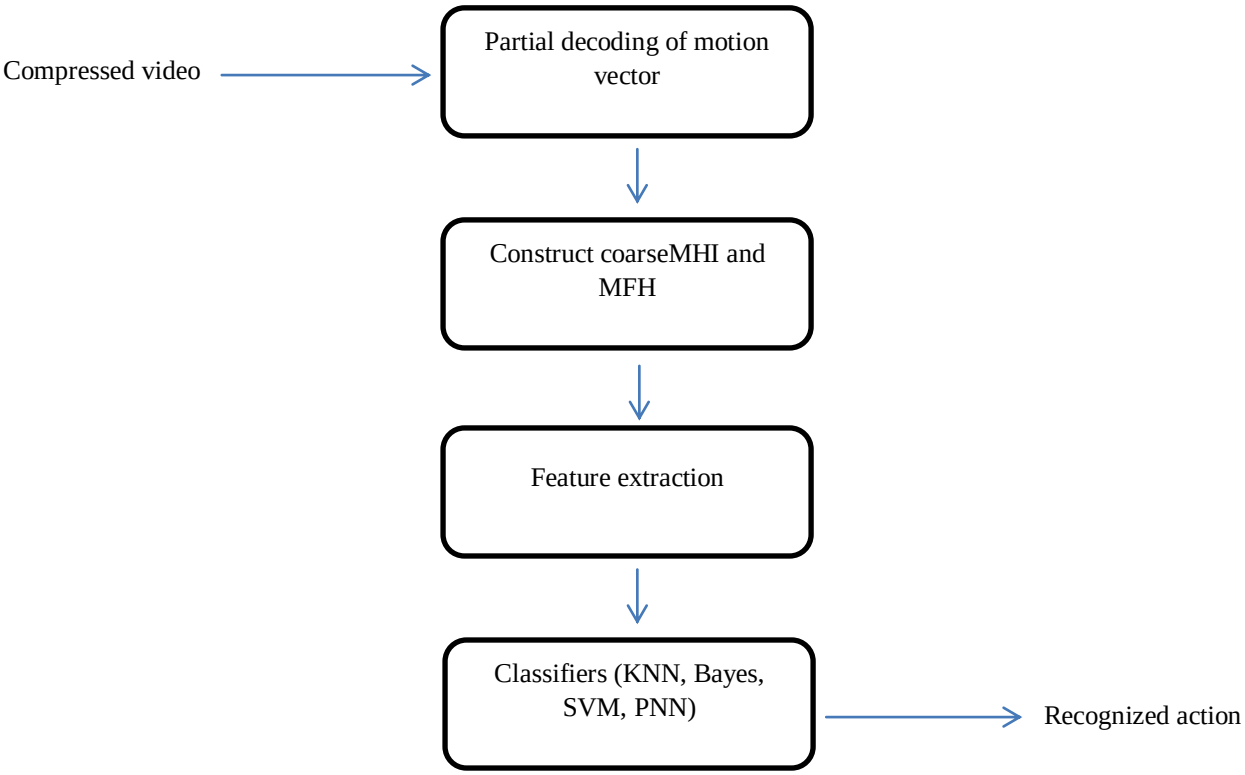


Figure:1

Firstly, the motion vectors are extracted from the compressed video by partially decoding the MPEG video bit – stream. This partial decoding is far less expensive as compare to the full decoding. As the motion vectors are usually noisy, the coarse MHI and MFH are constructed after removing the noisy motion vectors. The constructed coarse MHI and MFH are at macroblock resolution and not at pixel resolution. Hence the size of the MHI and MFH are sixteen times smaller than the original frame size i.e. $16^2$ times smaller in terms of number of pixels. In feature extraction phase, various features are extracted from the constructed coarse MHI and MFH, which hold the temporal and motion information of the video. The features are extracted from the MHI are projection profiles and centroids. The features are extracted from MFH are Affine features and Motion Vector Histogram. These features are finally fed to the classifiers such as KNN, Bayes, Neural Network, SVM and Probalistic Neural Network for recognizing the human action.

## III. FEATURE EXTRACTION FROM MHI AND MFH

MHI is the cumulative gray scale image incorporating the spatial as well as the temporal information of the motion. MHI points to, ‘ where and when did the motion occurs?’. It does not give any information about the direction and magnitude of the motion.

MFH gives the information about the extent of the motion at each macroblock. MFH points to, ‘where and how much did the motion occur?’

Extract some useful features for classification from MHI and MFH.

1 MHI features :

a) Projection profile based features: It indicates the bias of the MHI along horizontal and vertical direction w.r.t the centroid of MEI. This indirectly will conveys the temporal information of motion along horizontal and vertical direction.

b) Centroid based feature: The centroid of MHI and MEI is different because it is computed using grey-level time stamp values as weights in the summation. It gives the approximate direction of the movement of centroid for the corresponding action.

2. MFH features:

a) Affine feature: Though it is difficult to capture some complex motion affine model give a good approximation to the actual flow of the planar surface under orthographic projection. The affine parameters are estimated by standard linear regression techniques. The regression is applied separately on each motion vector component since the x affine parameter depends only on horizontal component of motion vector and y parameter depends only on vertical component of motion vector.

 

b) Projected 1D feature: The horizontal and vertical components of the motion vectors are considered separately. The histogram values are quantized into 5 bins to cover the entire range of the intervals. The bins are chosen in such a way so as to capture the low, medium and high speed.

c) 2D Polar feature: The angular direction and magnitude of motion vectors are considered together to quantize the polar plane into histogram bins. Each bin is defined by the angular range as well as magnitude (radius) range.

| | **Features** |
|---|---|
| MHI Based | Projection profile |
| | Centroid |
| MFH Based | Affine |
| | 1D Projected |
| | 1D Polar |

## IV. Conclusion

In this paper, we have proposed a method for constructing coarse MHI and MFHfrom compressed MPEG video with minimal decoding. Various useful features are extracted from the above mentioned motion representations for human action recognition. The performance of these features which is extracted from MHI and MFH is analyzed and compared. Though the test instances are from entirely different subjects other than those used for training he classifiers, the result show excellent human action recognition accuracy. The KNN. Neural Network (MLP) and SVM give the classification accuracy good and the Probalistic Neural Network give the best classification accuracy of 100% and 1D projected and 2D polar features show consistent performance with all the classifiers. Since the data is handled at macroblock level, the computational cost is extremely less compared to the pixel domain processing.

**References**

[1] J.K. Aggarwal, Q. Cai. Human motion analysis: a review, computer vision and image understanding 73 (3) (1999) 428-440.

[2] D. Ayers, M. Shah, Monitoring human behavior from videos taken in an office environment, Image and vision computing 19 (12-1) (2001) 833-846.

[3] R. venkatesh Babu, B. Ananthraman, K. R. Ramakrishnan, S.H. Srinivasan, Compressed domain action classification using HMM, Pattern Recognition Letters 23 (10) (2002) 1203-1213.

[4] R. Venkatesh Babu, K. R. Ramakrishnan, Compressed Domain Human Motion Recognition Using Motion History Information, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April, vol. 3, 2003, pp. 41-44

[5] A.F. Bobick, J.W. Bavis, The recognition of human movement using temporal templates, IEEE Transaction on pattern analysis and machine intelligence 23(3) (2001) 257-267.

[6] A.F. Bobick, A.D. Wilson, A state-based approach to the representation and recognition of gesture, IEEE Transactions on pattern analysis and machine intelligence 19 (12) (1997) 1325-1337.

[7] K. Boehm, W. Broll, M. Sokolewicz, Dynamic gesture recognition using neural networks; a fundamental for advanced interaction construction, Proceedings of the SPIE- The International Society for Optical Engineering 2177 (1994) 336-346.

[8] C. Bregler, Learning and recognizing human dynamics in video sequences, Proceedings of the IEEE CVPR (1997) 568-574.

[9] T. J. Darell, A.P. Pentland, Space-time gestures, Proceedings of IEEE Conference on Computer Vision and pattern recognition (1993) 335-340.

[10] J. Davis, Hierarchical motion history images for recognizing human motion, IEEE Workshop on detection and recognition of events in video (2001) 39-46.

[11] J.Davis, A.F. Bobick, The representation and recognition of human movements using temporal templates, Proceedings of the IEEE CVPR (1997) 928-934.

[12] D.J. Fleet, M.J. Black, Y. Yacoob, A. Jepson, Design and use of linear models for image motion analysis, International Journal of Computer Vision 36 (3) (2000) 171-193.

[13] D.M. Gavrila, The visual analysis of human movement: a survey, Computer Vision and Image Understanding 73 (1) (1999) 82-98.

[14] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, 1998.

[15] B.K.P. Horn, B.G. Schunck, Determining optical flow, Artificial Intelligence 17 (1981) 185-203.

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

 

[17] Y.A. Ivanov, A.F. Bobick, Recognition of visual activities and interactions by stochastic parsing, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 852-872.
[18] T. Kohonen, The self-organizing map, proceedings of the IEEE 78 (9) (1990) 1464-1480.
[19] M. Su, H. Huang, C. Lin, Application of Neural Networks in spatio temporal hand gesture recognition, Proceedings of the IEEE World congress on Computational Intelligence, 1998.
[20] A. Madabhushi, J.K. Aggarwal, A Bayesian Approach to Human Activity Recognition, Second IEEE Workshop on Visual Surveillance, 25-30, 1999.
[21] J.L. Mitchell, W.B. Pennebaker, C.E. Fogg, D.J. LeGall, MPEG Video Compression Standard, International Thomson Publishing, 1996.
[22] C.W. Ng., S. Ranganath, Real-time gesture recognition system and application, Image and Vision Computing 20 (13-14) (2002) 993-1007.
[23] N.M. Oliver, B. Rosario, A. pentland, A Bayesian Computer vision system for modeling human interactions, IEEE Transactions on pattern analysis and machine intelligemce 22 (8) (2000) 831-843.
[24] R.Polana, R. Nelson, Low level recognition of human motion, Workshop on Non-Rigid Motion (1994) 77-82.
[25] A. Psarrou, S. Gong, M. Walter, Recognition of human gestures and behavior based on motion trajectories, Image and Vision Computing 20 (5-6) (2002) 349-358.
[26] J. Rittscher, A. Blake, S.J. Roberts, Towards the automatic analysis of complex human body motions, Image and Vision Computing 20 (12) (2002) 905-916.
[27] R. Rosales, Recognition of human action based on moment based features, Technical Report BU 98-020, Boston University, Computer Science, 1998.
[28] M.Shah, R. Jain, Motion Based Recognition, Kluwer, 1997.
[29] Standard MPEGI: ISO/IEC 11172, coding of moving pictures and associated audio for digital storage media at upto about 1.5 Mbit/s, 1996.
[30] T. Starner, A. Pentland, Real-Time American Sign language recognition from video using hidden markov models, MIT Media Lab, TR-375, 1995.
[31] T. Starner, A.P. Pentland, Real-Time American Sign Language recognition using desk and wearable computer based video, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (12) (1998) 1371-1375.
[32] X. Sun, B.S. Manjunath, Panoramic capturing and recognition of human activity, IEEE International Conference on Image Processing 2 (2002) 813-816.
[33] M. Umeda, Recognition of multi-font printed chienese character, Proceedings of sixth computer vision and Pattern recognition (1982) 793-796.
[34] V.N. Vapnik, The Nature of Statistical Learning Theory, springer, New York, 1995.
[35] J. Yamato, J. Ohya, K. Ishii, Recognizing human action in time sequential images using hidden markov model, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (1992) 379-385.

 