International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869, Volume-2, Issue-5, May 2014

Survey on Human Action Recognition

Deepali kaushik

***Abstract*— Human activity recognition is a most interesting analysis among the computer vision and video processing community.it is an imperative area of computer vision research. Its application include patient monitoring system, variety of system that involve the interaction between human and object. In this review paper we term a system for recognition of human action with the help of different-2 techniques. Firstly we introduce the technique MHI and MFH for feature extraction from the compressed video and then these extracted features are used to train the KNN, Neural Network, SVM, Bayes classifiers for recognition the human action and secondly we will explain some other methodology for simple human action (low level) and high level activities such as single layer approach is used to recognize the simple human action recognition and hierarchical approach is used for high level action recognition discussed in this paper.**

***Index Terms*— MHI, MFH, KNN, SVM, Human activity recognition**

I. INTRODUCTION

Human activity recognition and feature extraction is an important area of computer vision research. To recognition the human action from the original video is problematic because it take more space and time. So firstly we compressed the video without any loss of information, it is simply convert the large space video into small space video. In the recent past, we reported a technique for human action recognition from the compressed video using Hidden Markov Model (HMM) [1]. The time series is used for training the HMM which is directed extracted. The extracted time features are not suitable for other efficient classifier such as k-nearest neighbour (KNN), Neural Network, SVM and Bayes.

In this review paper we offered a technique for building coarse motion history image (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, whereas MFH quantify the motion at the image plane. The feature extracted from MHI and MFH were used to train the KNN, Bayes, Neural Network, SVM classifiers for recognition the human action. These technique signify the human action in a very compacted manner. This work is motivated by a technique proposed by Davis and Bobick [2] where a view-based approach is used to recognize actions. They are presented a method for recognized of temporal templates. These techniques are used only for simple human based action recognition. The ability to identify complex human activities from videos define some important applications.

Manuscript received May 13, 2014.

Deepali kaushik, Department of computer science & engineering, Krishna institute of engineering & technology (KIET) Ghaziabad

There are various types of human activities depending on their complexity, we categorized human activity into 4 different levels such as gestures, actions, interactions and group activities. Gestures are elementary movement of a person's body part as 'stretching an arm' and 'raising a leg' are good example of gestures. Actions are activity of single person that is combination of multiple gestures such as 'walking', 'waving', 'punching'.The otheris, Interaction that are

the2 or more persons/ objects human activities for example- 'two person fighting', it is interaction between 2 humans [27]. Finally last activity is group activity which is performed by multiple persons/ objects. Ex- 2 group are fighting.

The previous review paper written by Aggarwal and Cai [13] has covered several essential low- level components for understanding of human motion such as tracking and body posture analysis. In this review paper we discuss on both level methodology designed for analysis of human actions. Fig-1 define the overview of the tree-structure taxonomy. All activity recognition methodology are first classified into 2 categories: single layered approach and hierarchical approaches.

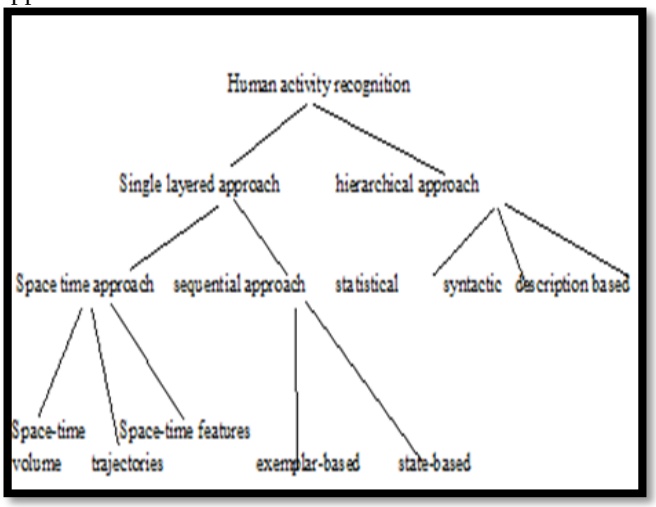


Fig. 1 The hierarchical approach based taxonomy of this review

Turaga et al. [14] is survey covered human activity recognition approaches however most of the previous review paper have focused on the introduction and summarization of activity recognition methodologies. In this review paper we present different technique for recognizing the human activity in broad area.

II. RELATED WORK

We will give a brief description of works associated to human motion and gesture recognition. To recognize the human activity in the low level we will defined into following two.

1) State- space based

2) Template matching based

State-space based approach: it uses the time series features obtained from the sequence of image. We will used HMM for activity recognition. It is done by yamato et al. [3]. The drawback of this method are that it is sensitive to position displacement, noise, poor performance if the training and test issue are different.

The gesture recognition work by Darrell and pentland [5] uses time- warping technique for recognition which is closely related to HMM.

There are few works reported in literature which use neural networks for gesture recognition [6,4]; boehm et al.[4] used kohenon feature maps (KFM) [7] for recognition dynamic gestures. Oliver et al. [8] proposed a system for modelling and recognizing human behaviour in a visual task.

Template matching based approach: one of the earlier work using this approach is found in the work done by polana and nelson [9] where the flow information is used as features. In this approach we compute the optical flow field [10] between successive frames and divide each frame into a spatial grid and find the sum. The motion magnitude to get the high dimensional feature davis&bobick [11,12] presented a real time approach for represent the human motion using compact MHI in pixel domain. The recognition performance user evaluated for the following 3 classifier namely KNN, Gaussian and mixture of the Gaussian.

III. METHODOLOGY

We follow this proposed system Fig. 2 to recognize the human action. This system is only used for the static representation as there is a method to capture & represent motion directlyfrom video that is MEI, MHI, and MFH.

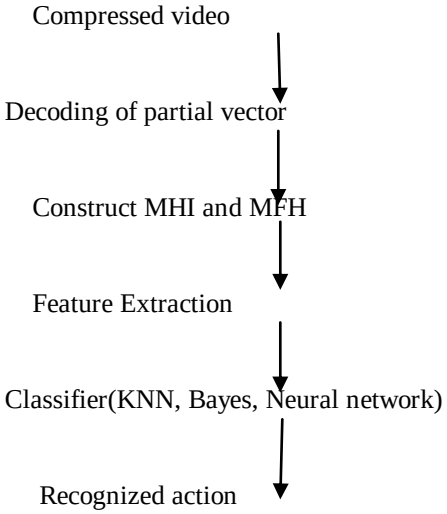


Fig. 2 Overview of proposed system

MEI represent the binary image with only spatial & no temporal details of the motion. MHI is a grey scale image [12]. That is point to where and when did the motion occur? It does not convey any information about the direction and magnitude of the motion. MFH gives the information about the size of the motion (where and how much did the motion occur?) and then we will extract the feature & then transfer into classifier. With the help of classifier the all action of human are recognized easily only for static parameterize. In Fig.3 MHI and MFH produce their respective result to recognition the features. As we define that MHI take only feature with the help of subtraction background from the original image and that shown in Fig 3a) and MFH define only direction of movement of extracted feature from the image that is shown in Fig 3b).

Another approach is defined above for high level human action recognized as-

Single layer approach recognized the human activity directly from video data. These application mainly used to be analyse the simple & sequential movement of human such as walking, jumping, and waving.

Single layer approach categorized into 2 classes- space time approach & sequential approach.

In space time approach, the system construct a model 3-D (XYT) space time volume to represent each activity. The video is sequence of 2-D image is formulated into 3D real world scene to analysis the human activity. So space time approaches are suitable for recognition periodic action and gestures. Space time approach further classified into 3 categories such as space time volume, space time trajectories and space time feature.

In **space time volume approach**, it provide a straight-forward solution but often have difficulties to handling speed and motion variation.

In **space time trajectories**, it is able to perform detailed-level analysis & it is view- invariant in the most cases.

Space time features approach is used to illumination changes & noise in the video. This approach is not suitable for modelling more complex activities [Niebles et al 23; Ryoo and Aggarwal 13].

Another approach for human activity recognition is **sequential approach**. Sequential approach is a single layer approach which is used to recognize the human activity by analysing the sequence of features. With the help of sequential approach, it is firstly convert the image into feature vectors and then describe the states of a person. Once the features vector have been extracted, then we can comparison with the original image feature with the high similarity.

Sequential approach classify into 2 Categories: exemplar-based recognition approach and state model based recognition approach.

Using the **exemplar- based approach**, the new input video compare with the sequence of feature vector which is extracted from the video with template sequence. If the similarity of feature vector is high enough then system is able to deduce that given input contain the activity. The dynamic time wraping (DTW) algorithm is widely used for matching the 2 sequence with variation [Darrell and Pentland 17; Gavrita and Davis 18].

The DTW algorithm finds an optimal non -linear match between 2 sequences and other approach for recognition of human activity is **state model- based approach**. It is also sequential approach to detect the activity to recognize the human activity. It composed a system to set of states which is statistically trained and match with feature vector to states.

These approach are only used for low-level activity which was defined in above.

Another approach is **hierarchical approach** which is applicable for high-level activity recognition as interaction and group activity etc.

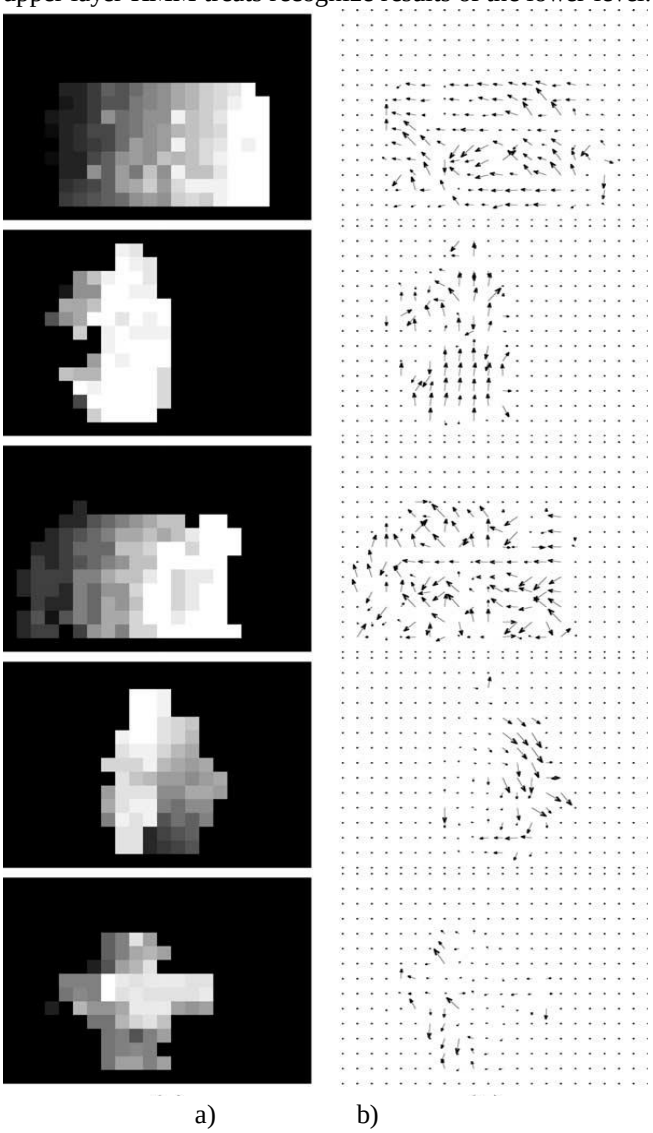
With the help of hierarchical approach, it can find multiple & multilevel human activity recognition. Hierarchical approach not only makes the recognition process computationally tractable and conceptually understandable but also reduce redundancy in the recognition process. The main advantage of hierarchical approach over no-hierarchical approach is their ability to recognize the high-level activities with more complex structure. Hierarchical approach are especially suitable for a semantic-level analysis of interaction between humans & objects as well as complex group activity.

Using approach based taxonomy, we categorize hierarchical approach into 3 groups.

Statistical approaches, syntactic approaches and description-based approaches.

In the case of hierarchical statistical approach, multiple layer of state based model such as HMM and DBN are used to recognized activities with sequential structure.

Oliver et al [24] presented layer hiddenMarkova model. In this approach the bottom layer HMM recognize the atomic action of a single person by matching the model with the sequence of feature vectors extracted from video and the upper layer HMM treats recognize results of the lower level.



a) b)

Fig. 3 (a) The coarse MHI and the corresponding (b) MFH of walk, jump, run,bend-up & twist-right action

Statistical approach we used when recognized sequential activities. If there is enough training data & noisy then it can easily recognize the activities. The major limitation of statistical approach are that unable to recognize activities with complex temporal structure.

Syntactic approach is used to recognition the action corresponds to an atomic level ivanov and bobic [20] proposed a hierarchical approach for recognition high-level activities using SCFG. They divide the framework into 2 layer. The lower layer using HMM for recognition of simple action and higher layer using stochastic parsing technique for the recognition of higher level activities. Syntactic approach are able to probabilistically recognize hierarchical activities composed of sequential sub events but are inherently limited on activities composed of concurrent sub-events.

Last approach for higher level recognition human action is description based approaches. Description based approach represent a high level human activity in terms of simpler activities composing the activities and describing their temporal, spatial and logical relationship. This approach model a human activity as an occurrence of sub- events. They use sub-events to represent human activities. In description based approach, a CFG is often used as syntax for representation of human activities [Nevita et al 21; Ryoo& Aggarwal 25, 26].

A Bayesian belief network is constructed for the recognition of the activity, based on its temporal structure representation. The root node of the belief network shows to high level activity that system aim to recognize.

IV. COMPARISON

Finally we can say that hierarchical approach are suitable for recognizing the high level activities which is decomposed into sub- events. [Oliver et al 24; Nevatia et al 22] statistical and syntactic approach provide a probabilistic framework for consistent recognition with noisy inputs. Whereas description based approach are able to represent the human activity with complex temporal structure. With the help of description based approach we can find not only sequential but also concurrent sub- events are handled.

The major drawback of description based approach are, that is unable to recognize the activity of low-level components (gestures detection failure) whereas single layer approach are used to only simple action recognition and using MHI and MFH that is able to extract the feature and further forward to classifier for recognition the action of human. So we can say that this approach is only applicable for single level action or the low level feature recognition. We can see the performance of classification accuracy using classifier to recognition of human action in table 1. This table show the performance only for simple human action recognition. As we discussed other method to recognition low-level and high-level action that are also show the performance, in table 2, table 3, table 4 the abilities of recognition shown using space time approach , sequential approach and hierarchical

approach. With the help of these table we can measure the abilities of their approach.

Classifier	No of feature used	Classification accuracy (%)
KNN	32	98.0
Neural Net	32	98.0
SVM	32	98.0
Bayes	4	94.1

Table. 1 Comparison of various classifiers

Approach type	Authors	Required low-levels	Structural consideration	Scale invariant	localization	View invariant	Multiple activities
Space- time volume	Bobick and J.Davis '01	Background	Volume-based	Templates needed	✓		
	Shechtman and Irani '05	None	Volume-based	Scaling required	✓		
	Ke et al. '07	None	Volume-based	Templates needed	✓		
	Rodriguez et al. '08	None	Volume-based	✓	✓		
Space- time trajectories	Campbell and Bobick '95	Body-part estimation		✓	✓	✓	
	Rao and Shah '01	Skin detection	Ordering only	✓	✓	✓	
	Sheikh et al. '05	Body-part estimation	Ordering only	✓	✓	✓	
Space-time feature	Chomat and Crowley '99	None	Ordering only	✓	✓		
	Zalnik-Manor and Irani '01	None		✓			
	Laptev and Lindeberg '03	None		✓	✓		
	Shuldt et al. '04	None		✓			
	Dollar et al. '05	None		✓			
	Yilmaz and Shah '05a	Background	Ordering only	✓	✓	✓	
	Blank et al. '05	Background		✓	✓		
	Niebles et al. '06	None		✓	✓		✓
	Wong et al. '07	None		✓	✓		
	Savarese et al. '08	None	Proximitybased	✓	✓		✓
	Liu and Shah '08	None	Co-occur only	✓			
	Laptev et al. '08	None	Grid-based	✓			
	Ryoo and Aggarwal '09b	None		✓	✓		✓

Table.2 comparing the abilities of the important space- time approach

Type	Approaches	Required low-level	Execution variation	Probabilistic	Target activities
Exemplar- based	Darrell and Pentland '93	None	✓ Linear only		Gesture-level
	Gavrila and L. Davis '95	Body-part estimation	✓		Gesture-level
	Yacoob and Black '98	Body-part estimation	✓		Gesture-level
	Efros et al. '03	Tracking	Linear only		Action –level
	Lublinerman et al. '06	Background subtraction	Linear only		Action –level
	Veeraraghavan et al. '06	Background subtraction	✓		Action –level
State model-based	Yamato et al. '92	Background subtraction	Model-based	✓	Action –level
	Starner and Pentland '95	Tracking	Model-based	✓	Gesture-level
	Bobick and Wilson '97	Tracking	Model-based	✓	Gesture-level
	Oliver et al. '00	Background subtraction	Model-based	✓	Interaction-level
	Park and Aggarwal '04	Background subtraction	Model-based	✓	Gesture-level

	Natarajan and Nevatia '07	Action recognition	Model-based	✓	Interaction-level
	Lv and Nevatia '07	3-D pose model	Model-based	✓	Action –level

Table 3 comparing among sequential approaches

Type	Approaches	Levels of hierarchy	Complex temporal relation	Complex logical concatenations	Recognition of recursive activities	Handle imperfect low-level
Statistical	Oliver et al. '02	limited (2-levels)				✓
	Shi et al. '04	limited (2-levels)	One relation 'before'			✓
	Damen and Hogg '09	limited (2-levels)				✓
Syntactic	Ivanov and Bobick'00	Unlimited	✓	✓	✓	✓
	Joo and Chellappa '06	Unlimited	✓	conjunctions only	✓	✓
Description- based	Pinhanez and Bobick '98	limited	network form only	network form only	✓	compensates 1 error
	Intille and Bobick '99	Unlimited	two relations	✓	✓	✓
	Siskind '01	Unlimited	sub-event	✓	✓	✓
	Ryoo and Aggarwal '09a	Unlimited	✓	✓	✓	✓

Table 4 comparing the abilities of the hierarchical approaches

V. CONCLUSION

In this review paper we have shown different 2 approaches for recognition the human action. For low-level action recognition we have used MHI, MFH from video that is used to extract the feature and then used classifier KNN, Neural network, SVM gives the best classification accuracy of 98% that is show consistent performance. This approach only used for low level action recognition and then we used single layer approach for gesture level action recognition and hierarchical approach for high level action recognition. In this review paper we have summarized the methodologies that have explored for recognition of human activities and discuss advantages and disadvantage of those approach. We have discuss non-hierarchical approach developed for the recognition of gesture and action as well as hierarchical approaches for the analysis of high level interaction between multiple human and object. Hierarchical approach have their advantage in recognition of high level activities performed by multiple person and they must be explored further in the future to support demands from surveillance system and other application.

REFERENCES

[1] R. VenkateshBabu, B. Anantharaman, K.R. Rama Krishnan , S.H.Srinivasan, Compressed domain action classification using HMM, Pattern Recognition Letters 23 (10) (2002) 1203–1213.
[2] J. Davis, Hierarchical motion history images for recognizing human motion, IEEE Workshop on Detection and Recognition of Events in Video (2001) 39–46.
[3] 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.
[4] K. Boehm, W. Broll, M. Sokolewicz, Dynamic gesture recognition using neural networks; a fundament for advanced interaction construction, Proceedings of the SPIE—The International Society for Optical Engineering 2177 (1994) 336–346.
[5] T.J. Darrell, A.P. Pentland, Space–time gestures, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (1993) 335–340.
[6] M. Su, H. Huang, C. Lin, C. Huang, C. Lin, Application of neural networks in Spatio temporal hand gesture recognition, Proceedings of the IEEE World Congress on Computational Intelligence, 1998.
[7] T. Kohonen, The self-organizing map, Proceedings of the IEEE 78 (9) (1990) 1464–1480.
[8] N.M. Oliver, B. Rosario, A. Pentland, A bayesian computer vision system for modeling human interactions, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 831–843.
[9] R. Polana, R. Nelson, Low level recognition of human motion,Workshop on Non-Rigid Motion (1994) 77-82.
[10] B.K.P. Horn, B.G. Schunck, Determining optical flow, Artificial Intelligence 17 (1981) 185–203.
[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] A.F. Bobick, J.W. Davis, The recognition of human movement using temporal templates, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (3) (2001) 257–267.
[13] Aggarwal, J. K. and Cai, Q. 1999. Human motion analysis: A review. Computer Vision and Image Understanding (CVIU) 73, 3, 428-440.
[14] Turaga, P., Chellappa, R., Subrahmanian, V. S., and Udrea, O. 2008. Machine recognition of human activities: A survey. IEEE Transactions on Circuits and Systems for Video Technology 18, 11 (Nov), 1473-1488.
[15] Bhargava, M., Chen, C.-C., Ryoo, M. S., and Aggarwal, J. K. 2007. Detection of abandoned objects in crowded environments. In IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS).
[16] Bregonzio, M., Gong, S., and Xiang, T. 2009. Recognising action as clouds of space-time interest points. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[17] Darrell, T. and Pentland, A. 1993. Space-time gestures. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 335-340.
[18] Gavrila, D. and Davis, L. 1995. Towards 3-D model-based tracking and recognition of human movement. In International Workshop on Face and Gesture Recognition. 272-277.
[19] Moore, D. J. and Essa, I. A. 2002. Recognizing multitasked activities from video using stochastic context-free grammar. In AAAI/IAAI. 770-776.
[20] Ivanov, Y. A. and Bobick, A. F. 2000. Recognition of visual activities and interactions by stochastic parsing. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 8, 852-872.
[21] Nevatia, R., Hobbs, J., and Bolles, B. 2004. An ontology for video event representation. In IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW). Vol. 7.
[22] Nevatia, R., Zhao, T., and Hongeng, S. 2003. Hierarchical language-based representation of events in video streams. In IEEE Workshop on Event Mining.

[23] Niebles, J. C., Wang, H., and Fei-Fei, L. 2006. Unsupervised learning of human action categories using spatial-temporal words. In British Machine Vision Conference (BMVC).

[24] Oliver, N., Horvitz, E., and Garg, A. 2002. Layered representations for human activity recognition. In IEEE International Conference on Multimodal Interfaces (ICMI). 3-8.

[25] Ryoo, M. S. and Aggarwal, J. K. 2009a. Semantic representation and recognition of continued and recursive human activities. International Journal of Computer Vision (IJCV) 32, 1, 1-24.

[26] Ryoo, M. S. and Aggarwal, J. K. 2009b. Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. In IEEE International Conference on Computer Vision (ICCV).

[27] J.K. Aggarwal and M.S Ryoo" human activity analysis. : a review" ACM journal.

Deepali kaushik, M.Tech (computer science) from KIET (Krishna institute of engineering & technology) Ghaziabad & B.Tech (computer science) from VIT Meerut.