SindhUniv. Res. Jour. (Sci. Ser.) Vol.50(003) 447- 452 (2018)

http://doi.org/10.26692/sujo/2018.09.0073

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Hierarchical structure for Modeling Human actions and Multi-Classifier Approach

N. MINALLAH[++], F. IMTIAZ, M. ASHFAQ

University of Engineering & Technology Peshawar, Pakistan

Received 2nd May 2018 and Revised 16th August 2018

Abstract-Automatic action recognitionis one of the challenging tasks in recent research. Complex human actions are composed of simpler motion patterns. This research relates to give a new direction for identifying human actions using hierarchical structure of motion patterns. The main contribution of our research is that we proposed hierarchical classifier based action recognition system. In past years, no analysis is done for recognition using different classifiers for each complex human actions with exploration of simpler motion patterns. We modeled our system by the fact that each human action originates from basic human movements. Human actions such as running and walking originate from legs movement and hand waving and hand clapping originate from hands movement. First different classifiers were used in our proposed single classifier hierarchical structure and then best performed classifier for each action is selected and applied in multi-classifier hierarchical structure. For classifier inputs, Spatio-Temporal Interest Points (STIP) are extracted using SIFT features from 50 consecutive frames of each action. Covariance of STIP features among action frames are used as feature vector for classification using KNN, SVM and Naïve Bayes. Hierarchical structure is implemented using single classifier approach where each classifier is used separately at each level of hierarchy. Analysis is done and it is concluded that each classifier behavior and performance is different for each action. Best classifiers are selected and integrated in hierarchical structure using multi-classifier approach. Results show that multi-classifier approach in hierarchical structure has improved results as compared to single classifier approach.

Keywords: Modeling Human Multi-Classifier Approach

1. INTRODUCTION

Visual surveillance is an important field of computer vision. Visual surveillance helps in monitoring surveillance situations like pedestrian monitoring, sport festivals, schools, colleges, streets, airports monitoring during peak hours. Availability of high resolution camera can provide detail visibility of small action in huge crowded areas. Due to demand of surveillance especially due to security risk, security threat and major terror attacks in cities and crowded place, automation of surveillance become utmost important. Automation of visual surveillance is one of top goals of researchers as intelligent visual surveillance system does not need active monitoring of human operator. In current situation any big city or mass transit system have large number of camera feeds to control room, so that it become a major challenge to monitor and alarm about specific event. Behavior analysis of human is more important in independent decision for monitoring. The accurate decision of behavior human and its correlation to its surrounding human behavior can judge a complete security situation. As due to digitalization of video stream, all camera feeds of surveillance are stored on digital taps. Detection of any specific moment from past events and correlating to current event are all depend on human operator. In case of live steam the feed are provided to different monitoring screen which monitored by human operator. Usually separate human operator is assigned for different section. Manual monitor of either live feed or stored video is creating intensive labor and can lead to error. Automation of accurate detection of surveillance missing of important events or behaviors due to human system can solve all such challenges. Different approaches have been adopted during studying its feature of video feeds. The approach of researcher varies and based on different factors like individual or group behavior finding, indoor or outdoor surveillance. Usually researcher analyzes the behavior in two major steps. Their first step includes the detection process and second step is basis on decision making. Researcher uses variable feature for detection of action. Detected action is classified with use of different classifiers.

2. RELATED WORK

(Qian 2010) used feature extraction using background subtraction and blog detection. Using energy motion and SVM multiclass classifier with binary tree architecture recognized the activity of human. (Chakraborty, and Bhaskar, 2012): used novel approach and selective STIP feature detection and SVM is used for classification. (Chen, 2016) proposed approach for human behavior detection based on limbs motion and moving body parts. Human behavior is detected with trained SVM classifier. used bag of words and features are extracted with SIFT combination with CNN. They used classifier SVM (linear, poly, additive che square kernel), KNN, random forestan and their combination with caffenet for further accuracy. (Christian and Ivan 2004) implemented complex motion

[++]Corresponding authors e-mail: n.minallah@uetpeshawar.edu.pk

pattern recognition using space time invariant features and SVM classification. (Ouanane and Serrier2013) proposed surf feature extraction and classification based on SVM. (Muhammad and Ganghu 2014) used action recognition using motion skeleton joint location. (Farzad and Niarakha2015) used recognition and classification based on hidden markov model. Our approach for detection of human based on hierarchical structure, implementation using different classifiers separately, analyzing and selecting best classifier at each level of hierarchy. After selection of best classifiers, multi-classifier hierarchical model is created. Hierarchical model with multi-classifier approach is new idea which is according to our knowledge not been implemented yet.

3. STIP FEATURES DETECTION

STIP function is selected for features extraction in action recognition. Spatio-temporal interest points based on spatio-temporal corners of the image.STIP features are extracted by constructing space scale using Gaussian. By calculating difference of guassian and locating extrema of its difference with subpixel localization, finds its location points.

For the image I(x,y) and variable scale Gaussian G(x,y,σ), the space scale L(x,y,σ)is obtained by convolution of image with variable scale gaussian.

$$\mathrm{L}(x,\ y,\sigma)\ =\ \mathrm{G}(x,\ y,\sigma)\ *\ \mathrm{I}(x,\ y) \tag{1}$$

where key points are located through difference of Gaussian. By difference of two images and locating scale space extrema, D(x,y,σ) can be calculated as:

$$\mathrm{D}(x,\ y,\sigma)\ =\ \mathrm{L}(x,\ y,\mathrm{k}\sigma)\ -\ \mathrm{L}(x,\ y,\sigma) \tag{2}$$

where k is the scale factor of one image with other. To find maxima and minima of D(x,y,σ), 8th neighbor comparison occur at same scale and 9th neighbor at one up and down scale. The max or min value at this point defines extrema. By laplacian key points are localized and low contrast images are removed.

$$\mathbf{z}\ =-\frac{\partial^2\mathrm{D}^{-1}}{\partial\mathrm{x}^2}\frac{\partial D}{\partial\mathrm{x}} \tag{3}$$

Where z is the location of extremum. The value of z is lower than threshold value, it will be removed. In our experimental setup top 40 extrema are selected among all features. The orientation of 40 extrema are calculated by further smoothing image through guassian. The magnitude m and direction θ can be calculated as,

$$\mathrm{m(x,y)}=\sqrt{\begin{matrix}\big(L(x+1,y)-L(x-1,y)\big)^2+\cdots\\ (L(x,y+1)-L(x,y-1))^2\end{matrix}}\theta(x,y) \tag{4}$$

$$\theta=\tan^{-1}\frac{L(x+1,y)-L(x-1,y)}{L(x,y+1)-L(x,y-1)} \tag{5}$$

Key point descriptors are line up its orientation and weighted by key point i.e 1.5* keyscale. By using 16 histogram, 4*4 grid each having 8 orientation, their main direction vector shows the feature vector. These top 40features and their vector direction help in finding action.

4. PROPOSED APPROACH

During last few decades, the research main focus to detect human actions is usually based on different feature extraction algorithms that applied on motion patterns and body skeleton motions. Our approach for micro-behavior detection isbased on SIFT features. Extraction of features for action recognition dependson a lot of factors like orientation, scaling of camera, motion of body and external environmental factors.STIP algorithm provides solution to the problem of temporal alignment and shows outstanding invariance to geometric transformation and is independent of orientation, scale, alignment, rotation and viewpoint of image. As SIFT features detected are local so they are independent of segmentation problem. SIFT have significant result in illumination variation and background clustering. Popular Haris features detector isused that can detect high intensity variation both space and time by using spatio temporal corner. STIP features are extracted from first 50 frames of all videos of KTH dataset for action recognition. Each action contains 50 frames and each frame contains top 40 SIFT features. SIFT features from 50 fames for each action clip is further processed using covariance of features among 50 frames. For a single clip of action we have a single covariance matrix. The main contribution of our research is the introduction of hierarchical modeling of human actions which recognize human actions by categorizing them into two main categories i.e., legs movement and hands movement. Based on these parent categories, human actions are refined by falling in one of these categories. Parent nodes are further subdivided as child nodes i.e., legs movement is sub-categorized as running and walking while hands movement is subcategorized as hand waving and hand clapping. Action cannot be recognized unless it reaches child node. Covariance matrix created for each action is divided as 70%training dataset and 30% testing dataset. SVM, KNN and Naïve Bayes classifiers are used for classification at each node of hierarchical model for action recognition. Figure 1 shows hierarchical model of our proposed methodology. The given hierarchical model is used in two forms. First form is hierarchical model-single classifier approach used in analysis stage where a single classifier is used one at a time. After analysis second form is hierarchical model-multiclassifier approach in which best classifier is

selected out of all classifiers at each level of hierarchy and on each branch.

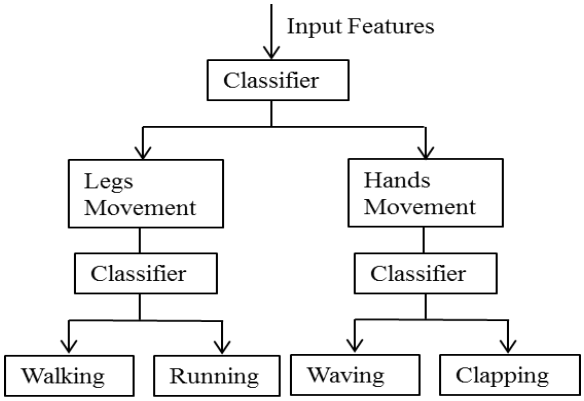

Fig. 1: Hierarchical model for action recognition

5. TECHNICAL DETAILS

Action recognition is one of the challenging tasks in field of computer vision. In this research we analyzed that each human action is originated from some basic human movements. Basic movements contain legs movement and hands movement. In this research we have taken four actions from KTH dataset i.e., walking, running, hand waving and hand clapping. All these four actions come under one of the basic body movements. In our research, we explored a new direction for recognizing human actions where each action can be identified by first recognizing the origination of these actions. Child categories are identified just after recognizing their parent categories. Classification is done by using KNN, SVM and Naïve Bayes classifiers, at each node of hierarchical structure for action recognition. Our research project is divided into different phases. In the first phase we have taken 50 frames out of each action from KTH action videos. Each action is mostly performed completely within 50 frames. In order to extract STIP features from 50 frames, SIFT features are taken. For action recognition only top 40 SIFT features are selected from each frame shown in figure 2. Each activity contains 50 frames with 40 features. These features can be directly taken as features vector but to make this process more effective we have generated covariance matrix for each activity. In past research covariance is taken among features within a single frame but not among consecutive frames of single activity. In this research, one more difference is that we have taken covariance among consecutive 50 frames of a single activity. These covariance matrices for each separate activity will then be used as input features to classification system for recognition. The feature set created is divided into 70% training and 30% of testing data for classification. Our proposed methodology contains hierarchical structure with legs movement and hands movements as the parent categories. We divided our dataset into two parts. All feature vectors for hand waving and hand clapping are grouped together to fall into category of hands movement. Similarly running and walking are grouped together to fall into category of legs movement. At parent level it is two category problem which to recognize whether it is legs or hands movement. Classifier is trained using KNN, SVM and Naïve Bayes separately. Once parent categories are recognized, we have two sub-categories of child. Separate classifier will be needed for each child category. Each classifier will be trained and tested in similar way as parent category. Each child level is also a two category problem as hand movement has two categories of hand waving and clapping while legs movement has two categories of walking and running. This hierarchical approach is applied first using by KNN on each level and then whole process is repeated using SVM and Naïve Bayes which we called it as hierarchical structure using single classifier approach. The main purpose of using different classifiers is to analyze behavior of a classifier for recognition of a specific action. Analysis is done using different classifiers separately and results show that each classifier's behavior is different for different actions in terms of accuracy. The last and final strategy is to integrate best classifiers at each level and even at each branch. After analysis, we used different classifier at each node which we called it as hierarchical structure using multi-classifier approach. Final results of integration of both classifiers are more accurate as compared to individual classifier.

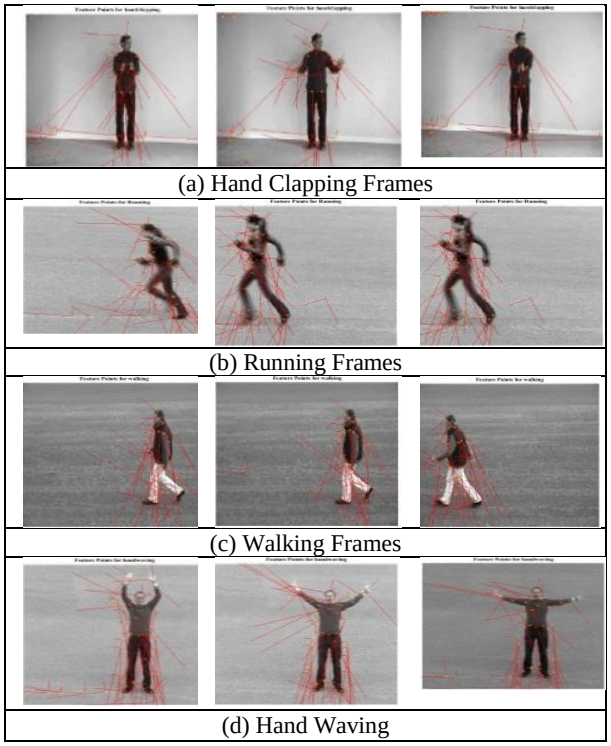

Fig.2: STIP features (magnitude and direction) shown by red lines of different actions frames

6. EXPERIMENTAL RESULTS

There are different phases of hierarchical structure for modeling action recognition. Results are generated first by designing and implementing hierarchical structure using SVM, KNN and Naïve Bayes separately or single classifier approach. Second phase is analysis phase in which results are compared and analyzed for best classifier for a specific hierarchical level. Third phase contains integration of these classifiers in a hierarchical model by selecting that classifier whose performance is better than other. Four actions are selected for our project from KTH dataset. Experiments are performed using 50 frames from each video. 50 frames almost complete one action. We create covariance matrix of a video considering 50 frames with 40 features each. Covariance is check among 50 frames of a video. So for 400 videos (100 video for each action) 400 covariance matrices are generated. These covariance matrices are further sub-divided as 70% training dataset and 30%testing dataset. Two main categories are created (Hands Movement and Legs Movement) so that we have the hierarchal model. Hands movement is further subdivided into two child categories i.e., hand waving and hand clapping. Legs movement is further subdivided into two child categories i.e., walking and running. Hand waving and clapping is the subset of the training dataset for hand movement and walking and running is the subset of training dataset of legs movement. Results are generated at each node of hierarchical structure. Details of accuracy using SVM, KNN and Naïve Bayes at parent level and child level are given in **(Table 1 and 2)** respectively which is done using single classifier approach. Results show that KNN accuracy is better than SVM and Naïve Bayes so it can be selected at parent level. At node child 1 SVM result is much better than KNN and Naïve Bayes so it can be selected at this node while at child 2 KNN or SVM can be selected as shown in table 3. It is to be noted that at child level overall accuracy of both classifier may be close to each other but accuracy for correct identification for specificaction may be different. So selection of classifier at this level can change according to requirement that which action is more important to be identified correctly.**(Fig. 3 and 4)** shows summary of all results of hierarchical structure using single classifier approach and multiclassifier approach respectively.

Table 1: Accuracy using KNN, SVM and Naïve Bayes using single classifier at Parent Level

Classifier	Category Type	Total no. of videos	Correct identifications	Accuracy for each action	Classifier accuracy
KNN (Parent)	Legs movement	60	58	96.7%	96.7%
	Hands movement	60	58	96.7%	
SVM (Parent)	Legs movement	60	58	96.7%	95%
	Hands movement	60	56	93.3%	
Naïve Bayes (Parent)	Legs movement	60	60	100%	71.7%
	Hands movement	60	26	43.3%	

Table 2: Accuracy using KNN, SVM and Naïve Bayes using single classifier at Child Level

Classifier	Category Type	Total no. of videos	Correct identifications	Accuracy for each action	Classifier accuracy
KNN (Child 1)	Walking	30	26	86.7%	73.3%
	Running	30	18	60%	
KNN(Child 2)	Hand Waving	30	13	43.3%	51.7%
	Hand Clapping	30	18	60%	
SVM (Child 1)	Walking	30	30	100%	75%
	Running	30	15	50%	
SVM (Child 2)	Hand Waving	30	24	80%	51.6%
	Hand Clapping	30	7	23.3%	
Naïve Bayes (Child 1)	Walking	30	1	3.3%	50%
	Running	30	29	96.7%	
Naïve Bayes (Child 2)	Hand Waving	30	30	100%	50%
	Hand Clapping	30	0	0%	

Table 3: Results of Hierarchical Structure using multi-classifier approach

Classifier Selected	Actions	Total no. of videos	Correct identifications	Accuracy for each action	Classifier accuracy
KNN (Parent)	Legs Movement	60	58	96.7%	96.7%
	Hands Movement	60	58	96.7%	
SVM (Child 1)	Walking	30	30	100%	75%
	Running	30	15	50%	
SVM (Child 2)	Hand Waving	30	24	80%	51.6%
	Hand Clapping	30	7	23.3%	

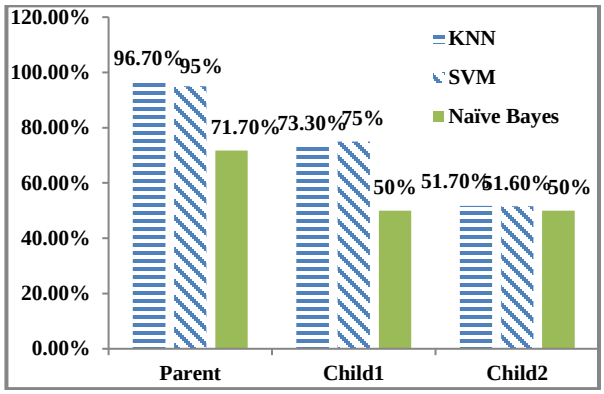

Fig. 3: Results of Hierarchical structure using single classifier approach

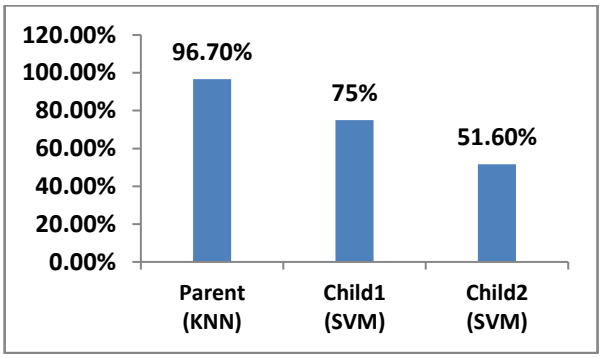

Fig. 4: Results of Hierarchical structure using multi-classifier approach

7. CONCLUSION

Human actions are composed of simple motion patterns. In this research we modeled hierarchical structure for modeling human actions that originate from simple body movements. Human actions are taken from KTH dataset. Legs movement and hand movement is placed at top level of hierarchy.Hierarchical structure is first modeled (Fig. 4):Results of Hierarchical structure using multi-classifier approachwith single classifier approach using different classifiers separately. Each classifier has different performance and accuracy for each action classification. Best classifier is selected at each level of hierarchy to implement multi-classifier approach. Results show that using single classifier may not be sufficient, multi-classifier approach improves the results. In future, more efficient techniques can be implemented in hierarchical structure using multi-classifier approach to further improve the results. Furthermore, research will be useful for modeling more complex human actions by categorizing in simple human body movements and recognition using multiclassifier hierarchical approach.

REFERENCES:

Bagheri, M. A, (2014)"A framework of multi-classifier fusion for human action recognition." *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE,

Chakraborty, and Bhaskar, (2012): "Selective spatio-temporal interest points." *Computer Vision and Image Understanding* 116.3 396-410.

Chen, Y. (2016) "Human Behavior Recognition Method based on Image Sequences." *International Journal of Signal Processing, Image Processing and Pattern Recognition* 9.2: 189-202.

Farzad, A. and R. N. Asli. (2015)"Recognition and classification of human behavior in Intelligent surveillance systems using Hidden Markov Model" *International Journal of Image, Graphics and Signal Processing* 7.12: 31.

Qian, H., (2010)"Recognition of human activities using SVM multi-class classifier." *Pattern Recognition Letters* 31.2: 100-111.

Ouanane, A., A. Serir, and N. Djelal. (2013) "Recognition of aggressive human behavior based on SURF and SVM." *Systems, Signal Processing and their Applications (WoSSPA), 2013 8th International Workshop on*. IEEE,

Schuldt, C., I. Laptev, and B. Caputo. (2004) "Recognizing human actions: a local SVM approach." *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. Vol. 3. IEEE,