International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 7, July 2017)

Human Activity Recognition Using HOG Features

Neha Aher[1], Milind Kamble[2]

***Abstract*— Human activity recognition is delivered information about the identity of a person, psychological state, and their personality. The human ability to recognize human activities is one of the leading subjects of study of the scientific areas of machine learning and computer vision. This paper presents a method to automatically identify human activity from input video stream using Histogram of Oriented Gradient features (HOG) and support vector machine (SVM) classifier was used for classification. Histograms of Oriented Gradient of differential images were taken; Differential images were obtained by taking the frame difference of successive frames in a video. Experiment was conducted on KTH, Weizmann and Kitchen activity database and gives performance with 72%, 85% and 85.4% accuracy respectively for test dataset. The experimental result demonstrates that our approach achieved a better and more stable performance.**

***Keywords*— Human Activity Recognition, Differential images, Feature Extraction, Histogram Oriented Gradient (HOG), Support Vector Machine (SVM).**

I. Introduction

Human activity recognition plays a vital role in interpersonal relations and human interactions. It provides information about the identity of a person, psychological state and their personality. One of the leading subjects of study of the scientific areas of computer vision and machine learning is the human ability to recognize human activities. Different application based on recognition of complex human activity from sequence of images i.e. videos. Detection of abnormal and suspicious activities at the airports, railway stations, shopping mall, military installations, crowded sports arenas are required [7]. Monitoring of elderly persons in smart health care facility, patients, children's are done by human activities recognition [8]. To capture and monitor scenes using closed circuit television (CCTV) by humans has become ubiquitous [7].

In activity recognition system, detection of human-being in the frames is the first step of human activity recognition.

One of the most standard and effective "person detectors" systems are used the HOG with SVM approach. In this paper, proposed activity recognition system tried to assess the qualitative characteristics of the human activity. For this purpose, the proposed activity recognition system first calculates HOG features of frames extracted from videos.

After extracting HOG features of all videos, some of videos are used for training and other are used for testing purpose using support vector machine classifier (SVM). The paper is organized as follows. In section 2 presents the literature survey and in section 3 covers detailed methodology consisting of three major steps, pre-processing, features extraction and classification. Section 4 experimental results. Section 5 concludes the paper.

II. Literature Survey

Video is a sequence of images and action is set of sequence of small movements. Human detection is first task of our algorithm so Navneet Dalal and Bill Triggs [9] are normalized descriptor blocks as Histogram of Oriented Gradient (HOG) descriptors. Tiling the detection window with a dense overlapping grid of HOG descriptors and used SVM for human detection. Wei-Lwun Lu, James J. Little [10] proposed HOG descriptor which was constructed by converting the tracking region to the grids of Histograms of Oriented Gradient (HOG) descriptor, and then used Principal Components Analysis (PCA) to project the HOG descriptor to a linear subspace. For recognition purpose Maximum Likelihood Estimation (MLE) was executed based on the observations using the Hidden Markov Model classifier. William Brendel and Sinisa Todorovic [11] presented an exemplar-based approach for identification of human postures where HOG features were computed and dictionary of discriminative features was learned. The videos were represented as temporal sequences of the learned code-words and activities in query video were detected by aligning the query and exemplar time series. Huang and Hsieh [12] recommended Histogram of motion history image (MHIHOG) for action recognition. It modernises an action sequence into a motion history image by collecting frame variance of the sequence. The motion history image is interpreted into a HOG descriptor, and then SVM classifier is used for categorize the actions.

Dipankar Das [13] proposed system which was divided into two phases, training and testing. In training stage the histogram of oriented gradient features (HOG) was extracted from each frame of the video. HOG feature vectors were extracted to generate the histogram of oriented gradient pattern history (HOGPH). HOGPH vectors were used to train multi-class SVM classifier model.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 7, July 2017)

In testing phase, HOGPH feature vector for each human activity was generated by the system and it was given to the trained SVM classifier for classification. This experiment proved that liblinear kernel produced higher accuracy than rbf kernel. Pooja G, Revansiddappa S Kinagi also recommended same approach for abnormal activities in [3].

Sabanadesan Umakanthan and Simon Denman, et al. [14] proposed effective method for feature representation where HOG and histogram of optical flow (HOF) from patch of densely sampled video at different scale were extracted and mi-SVM and K-mean clustering classifiers were used for multiple instance dictionary learning and codebooks for each class were built. Vaibhav Janbandhu [4] focused on classification of nonlinear human posture using Linear SVM classifier in which R-HOG (Rectangular) or C-HOG (circular) was used. Contrast normalization over overlapping spatial block was used for invariance of condition of awareness and shadowing. DoHyung Kim and Woo-han et al. [15] constructed a model which represented global shape of Depth Motion Appearance (DMA). Compact and discriminative actions were represented using HOG which were extracted from the DMAs and Depth Motion History (DMH).

Joshi, Shah [5] projected efficient method for human action detection using HOG features in which XML file of HOG features of positive and negative images of humans was generated and cascade classifier was trained. After the detection of human-being, ROI (Region of interest) part in the images were found and thresholding was used for recognition of object and activity. A. Jeyanthi Suresh, P.Asha [6] recommended system which was divided into two phases, training and testing. In training phase, HOG feature vectors were extracted from n consecutive video frames and were processed to generate action pattern. This action pattern was given to the Probabilistic Neural network (PNN) classifier for classification. In testing stage, for each human activity the action pattern was generated and was fed to the PNN classifier for action detection. PCA

was used for dimensionality reduction. K.P.Sanal Kumar and Dr.R.Bhavani [16] recommended combined feature for recognition of human activities. Histogram of Oriented Gradients (HOG), Motion Boundary Histogram (MBH) and Trajectory was fused together to create a single feature. Principal Component Analysis (PCA) was used for feature reduction and SVMkNN combined classifier was used for classification.

Florian Baumann [19] proposed effective framework using HOG and optical flow. Frame-by-frame learning approach was used to build two Random Forest classifiers independently and the ultimate conclusion was determined by combining both class probabilities. By using both features it eliminated illumination, contrast and background difficulties. Xiaodong Yang, Chenyang Zhang et al. [17] proposed method by using sequence of depth maps in which HOG features were extracted from the Depth Motion Maps (DMM). Compact and discriminative human actions were represented from any of the side views. Navneet Dalal et al. [18] introduced and estimated a figure of motion-based feature sets for human detection in videos. The detectors combined gradient based appearance descriptors with differential optical flow and constructed motion descriptors (IMHcd or IMHmd) in a linear SVM framework. Both motion and appearance channels used oriented histogram voting to achieve a robust descriptor.

III. METHODOLOGY

Based on the literature survey, HOG with SVM approach is one of the most widespread and effective "person detectors" used today. In this approach first extract one by one frame from video after certain time and convert into grey scale images. After that difference between consecutivee grey images is calculated and HOG is applied on that images. HOG features of each images of video are added and create the database of HOG features. Maximum number of dataset is used for training and remaining of it used for testing purpose. Support vector machine is trained, and after that test dataset given to that for classification as per figure 1.

A. Pre-processing

Based on the literature survey, HOG features with SVM approach is one of the most widespread and effective "person detector method" used today. In this approach first extract one by one frame from video after certain time and take difference and those difference images are convert into grey scale images as shown in figure 2

As shown in figure 2 taking two consecutive frames and third image is difference gray-scale image of these two images.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 7, July 2017)

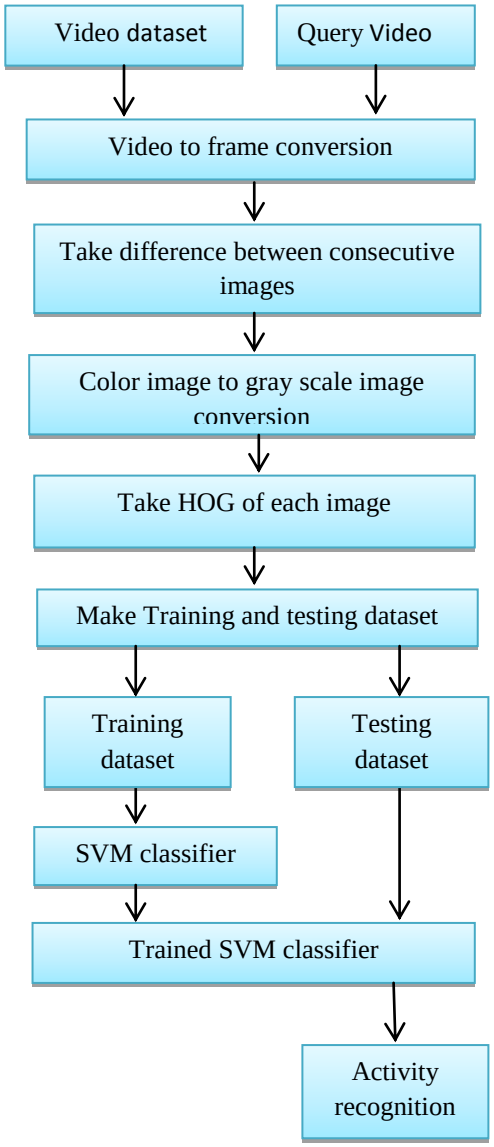
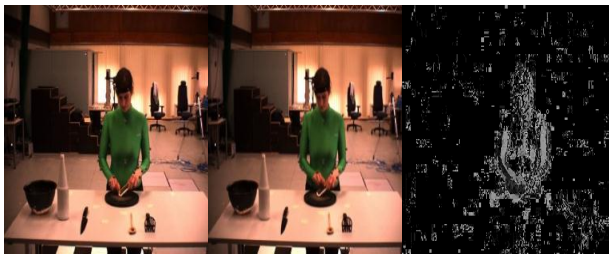


Fig.1. Block diagram of human activity recognition



a b c

Fig.2. (a) and (b) consecutive images (c) difference gray scale image

B. Histogram Oriented Gradient (HOG)

HOG is a type of "feature descriptor". The reason behind using HOG features is that instead of a "local" feature, it uses a "global" feature to describe a person [1]. The HOG uses a sliding detection window which is moved around the image. A HOG descriptor is calculated at each position of the image. This descriptor is then used for to make training set and testing set, training set feed to SVM classifier and trained SVM classifier categorized it as "person doing this activity or that activity". To calculate the HOG descriptor, it operated on 16x16 pixel cells within the detection window. Figure 3 (a) and (b) shows operator operated on images in x and y direction. These cells are organized into overlapping blocks. Algorithm takes 256 gradient vectors (in our 16x16 pixel cell) and a 9-bin histogram is made. The Histogram range starts from 0 to 180 degrees, so that there are 20 degrees per bin.

HOG feature is extracted from the gray-scale image applied as an input. These features are returned in 1-by-N vector, where N is the HOG feature length, which are used to encode local shape information from regions within an image as shown in figure 4.

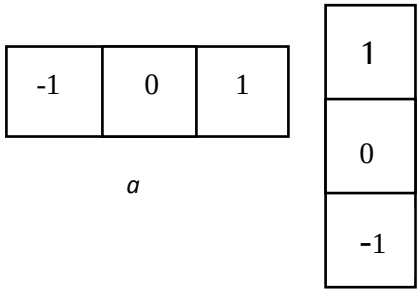
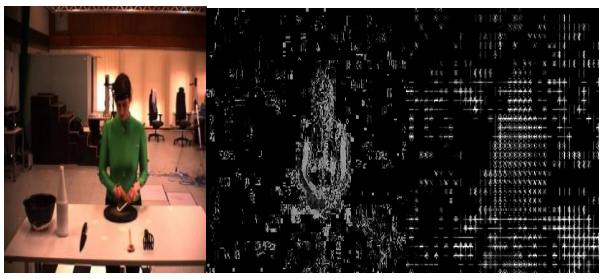


Fig.3. (a) Horizontal operator and (b) vertical operator



a b c

Fig.4. (a) original dataset (b) Gray-scale image (c) Image after applied HOG

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 7, July 2017)

(i.) Computing gradient of image: A gradient vector can be computed for each pixel an image. It's simply a measure of the change in pixel values along the x-direction and the y-direction around each pixel.
The equations for the magnitude and angle of a vector such as

$$Magnitude = \sqrt{(x)^2 + (y)^2} \quad (1)$$

$$Angle = arctan\left(\frac{y}{x}\right) \quad (2)$$

Where, x is the difference between two neighbouring pixels in horizontal direction and y is the difference between two neighbouring pixels in vertical direction.

Equation (1) and (2) used for computation of the gradient vector. The direction of the gradient vector is perpendicular to the edge is an important property of gradient vectors [2].

C. Support Vector Machine (SVM)

In machine learning, support vector machines are associated with learning algorithms for analysing and classifying data. Support vector machines are the potential mechanism for pattern classification problem. SVM is a bilinear classifier and maximize the decision boundary to reach the maximum separation between the object classes. Both linear and non-linear SVMs [3] are used for human detection in which non-linear SVM requires more computational cost.

In figure 5, straight line which separates two different classes is called border. The three blue filled triangle and two red filled circles are Support Vectors and the lines separating these classes are called Support Planes. Support planes are adequate for finding our decision boundary.

The equation of SVM classifier $\omega^T x + b_0$ is greater than 1 represent one class and less than -1 represent another class. In the equation w is weight vector (ω= [ω1, ω2,…… ωn]), x is feature vector (x=[x1,x2,……xn]) and b0 is bias. Here HOG is used as feature vector. Weight vector and bias point b0 decides the orientation of decision boundary and its position respectively.

Now decision boundary is defined to be midway between these hyper-planes so expressed as

$$\omega^T x + b_0 = 0 \quad (3)$$

The minimum distance between support vector and decision boundary is given by,

$$\text{Distance support vector} = \frac{1}{\|\omega\|} \quad (4)$$

Margin is twice this distance. It is maximized by minimizing a new function

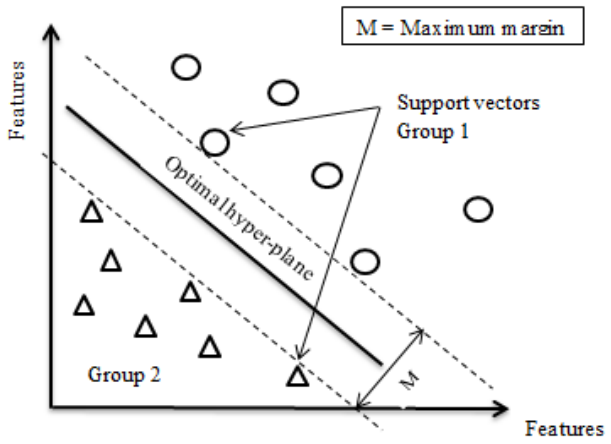


Fig.5. SVM hyper plain

$L(\omega, b_0)$ with some constraints [20] which are as follows:

$$\min L(\omega, b_0) = \frac{1}{2}\|\omega\|^2 \text{ subject to } t_i(\omega^T x + b_0) \geq 1 \forall i \quad (4)$$

IV. Experimental Result

In this algorithm, first differences between consecutive images are obtained and HOG is computed on all differential images in each video. KTH, Weizmann and Kitchen datasets are used for analysis and trained Support Vector Machine classifier. Figure 6 illustrates that as numbers of training samples increases, accuracy of algorithm also increases so that kept more number of videos for train support vector machine. All dataset are examined keeping more number of training samples compare with testing samples.

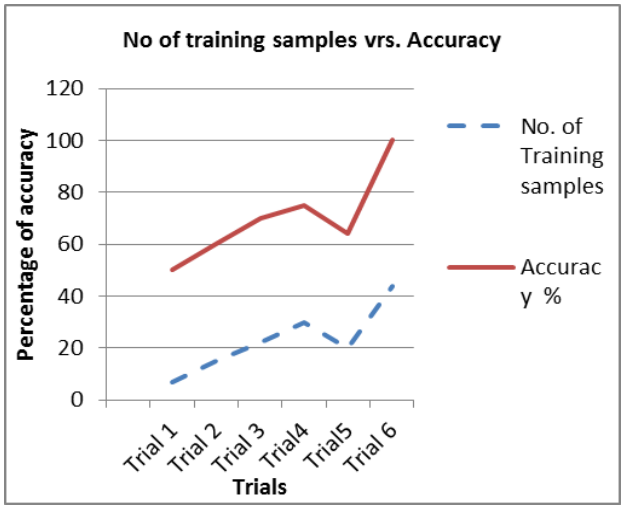


Fig.6. Analysis of number of training samples verses accuracy

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 7, July 2017)

(i.) KTH Dataset: The KTH is a well-recognized publicly available dataset for single human action recognition. The dataset for single human action recognition consists of 600 video files from 25 subjects performing six actions (boxing, clapping, walking, jogging, running and waving) in 4 different scenarios. Different people perform the same action at different directions and speeds. Figure 7 shows sample action images from KTH activity data set. The four different scenarios such as D1,D2,D3 and D4 where D1 is outdoor, D2 is zoom in and zoom out means different scales, D3 is outdoors and human wearing different cloths D4 is indoor.

On KTH dataset, HOG is computed on difference frames in video as explained in section 3.2. The dimension of HOG feature vector is of size 1x560 each and sum of all HOG features of frames in one video are taken; simultaneously this procedure is followed for each video.

Figure 8 provides experimental results for each action in the dataset under different circumstances. At time two actions are compared; D1, D2, D3 and D4 are the different circumstances. Hand clapping and hand waving actions are classified with 100% accuracy in D2 and D3 scenario. Average correct recognition rate of 72% is achieved for all actions in KTH dataset.

a b c d

Fig.7. (a) and (b) consecutive KTH dataset images (c) Difference gray image (d) HOG image

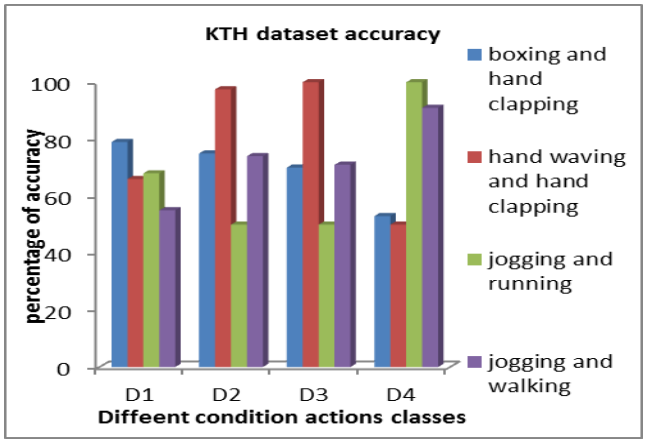
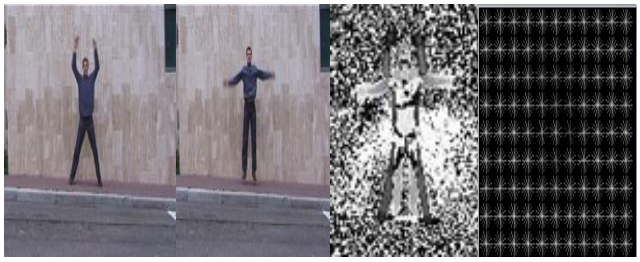


Fig.8. Accuracy of KTH dataset in different conditions

(ii.) Weizmann dataset: In 2005, the Weizmann Actions Dataset was recorded with an aim of studying new algorithms that could improve the human action recognition systems present at that time. The background is quite straightforward and only one person is acting in each frame. It contains 10 human actions which is a set of 9 actions: jump, bend, run, walk, jumping jack, gallop sideways, skip, one-hand wave, two-hand wave. Figure 9 shows sample action images from Weizmann activity data set. Two actions are compared at a time. In Weizmann dataset, HOG is computed on frames in video as explained in section 3.2.The dimension of HOG feature vectors is of size 1x792 each and the sum of all HOG features of frames are taken in one video.

Figure 10 provides experimental results for each action in the dataset. At time two actions are compared; jump and bend, run and walk, jack and side actions are classified with 100% accuracy. Average correct recognition rate of 85% is achieved for all actions in Weizmann dataset.



a b c d

Fig.9. (a) and (b) consecutive Weizmann dataset images (c) Difference gray image (d) HOG image

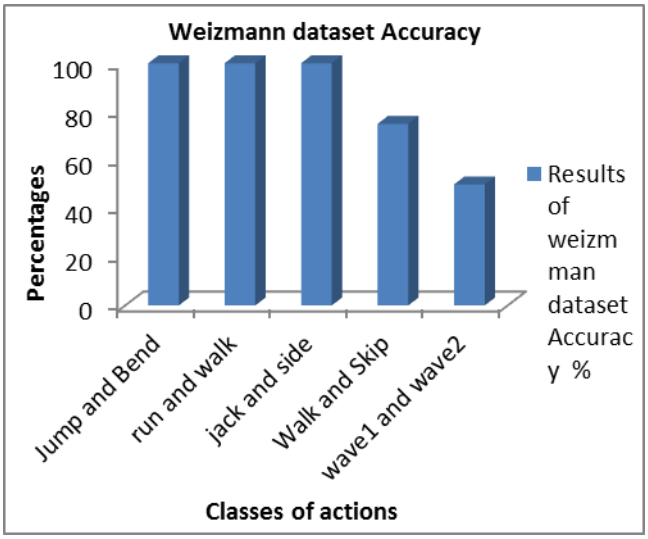


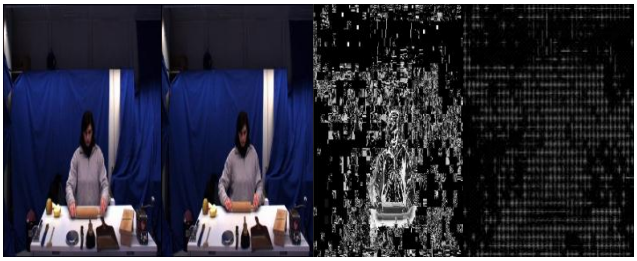
Fig.10. Accuracy of weizmann dataset

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 7, July 2017)

(iii.) Kitchen dataset: The background is quite simple and only one person is acting in each frame. It contains two human actions which are a set of 10 actions: mill, rollout, pour, chopping, saw, slice, stir, swap, grate and mash. In kitchen dataset two humans are doing actions wearing different cloths. Example frames from this dataset are shown in Figure 11.

Figure 12 provides experimental results for each action in the dataset. At time two actions are compared; mill and rollout, stir and swap actions are classified with 100% accuracy. Average correct recognition rate of 85.4% is achieved for all actions in Kitchen dataset.

Table I provides comparative study between proposed method with other methods used by people and proposed method gives better results compare to others.



A b c d

Fig.11. (a) and (b) consecutive Kitchen dataset images (c) Difference gray image (d) HOG image

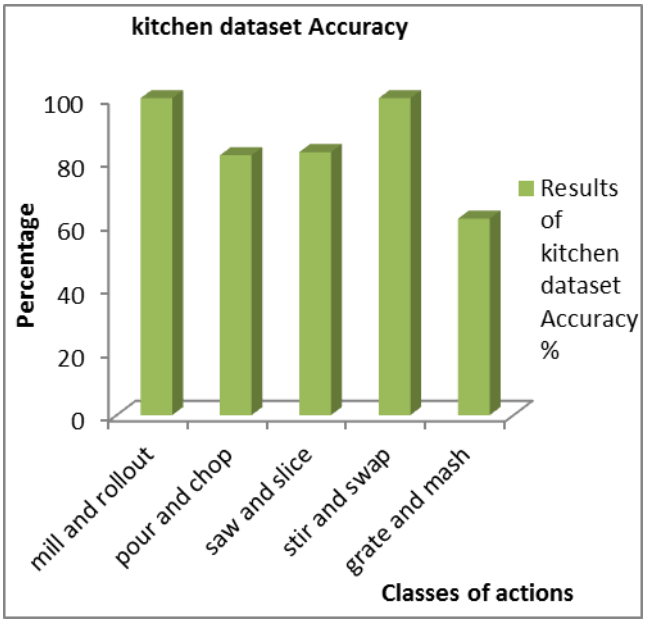


Fig.12. Accuracy of kitchen dataset

TABLE I
COMPARISON OF DIFFERENT METHODOLOGIES

Method+classifier	Dataset used	Recognition Rate (%)	No. of activities
MultiSVM+HOG	Willow action	61.07	6
DCDS+SVM	Weizmann	84	10
HOG+SVM	KTH	88.1	6
HOG+SVMkNN	Dataset used	72.48	4
Proposed method	KTH Weizmann Kitchen	73 85 86	8 10 10

V. CONCLUSION AND FUTURE WORK

In this paper, instinctive human activity recognition system is recommended that accurately recognizes different activities from real-life video data. The proposed system uses the HOG feature vectors that represent human activity and its behavior accurately and precisely. The HOG feature vectors are used to learn and classify human activity using SVM classifier with linear kernel. The proposed system is evaluated using validation strategies on the well-known, publicly accessible KTH dataset, Weizmann dataset and kitchen dataset and gives 72%, 85% and 85.4% accuracy respectively for single human action recognition.

Datasets used for analysis are recorded using steady cameras, any movement of camera adds noise in frames and increases missed classification. Future work will concentrate on extending proposed method in order to eliminate noise due to moving cameras in human activity recognition.

REFERENCES

[1] 'HOG person detector', http://mccormickml.com/2013/05/09/hog-person-detector-tutorial

[2] 'Gradient vectors', http://mccormickml.com/2013/05/07/ gradient-vectors

[3] Pooja G, Revansiddappa S Kinagi.: "Abnormal Activity Detection using HOG Features and SVM Classifier", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 5, Issue 4, April 2016.

[4] Vaibhav Janbandhu.: "Human Detection with Non Linear Classification Using Linear SVM", Volume 3 Issue 12, December 2014.

[5] Aarsh Joshi, Chandni Shah, Khyati Jain, et.al.: "A Review on Human Activity Recognition using HOG Feature", International Journal for Scientific Research & Development| Vol. 3, Issue 11, 2016 | ISSN (online): 2321-0613.

[6] A. Jeyanthi Suresh, P.Asha.: "Human Action Recognition in Video using Histogram of Oriented Gradient (HOG) Features and Probabilistic Neural Network (PNN)", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 7, July 2016.

[7] Daniel Weinland, Remi Ronfard, Edmond Boyer.: "Free viewpoint action recognition using motion history volumes", computer vision image understanding of a person's shape, appearance, movement, and behaviour Vol.104 November 2006, pp.249-257.

[8] Aaron F. Bobick, James W. Davis.: "The Recognition of Human Movement Using Temporal Templates", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 23, no. 3, march 2001.

[9] Navneet Dalal and Bill Triggs.: "Histograms of Oriented Gradients for Human Detection", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)

[10] Wei-Lwun Lu, James J. Little.: "Simultaneous Tracking and Action Recognition using the PCA-HOG Descriptor", The 3rd Canadian IEEE Conference on Computer and Robot Vision (CRV'06)

[11] William Brendel and Sinisa Todorovic.: "Activities as Time Series of Human Postures", 11th European Conference on Computer Vision, Crete, Greece, 2010

[12] Chin-Pan Huang, Chaur-Heh Hsieh, Kuan-Ting Lai*, et.al: "Human Action Recognition Using Histogram of Oriented Gradient of Motion History Image", International Conference on Instrumentation, Measurement, Computer, Communication and Control, 2011.

[13] Dipankar Das.: "Activity Recognition Using Histogram Of Oriented Gradient Pattern History", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol. 4, No.4, August 2014.

[14] Sabanadesan Umakanthan, Simon Denman, Clinton Fookes and Sridha Sridharan.: "Multiple Instance Dictionary Learning for Activity Representation", 22nd International Conference on Pattern Recognition, 2014.

[15] DoHyung Kim, Woo-han Yun, Ho-Sub Yoon, et.al: "Action Recognition with Depth Maps Using HOG Descriptors of Multi-view Motion Appearance and History", The Eighth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies 2014.

[16] K.P.Sanal Kumar, Dr.R.Bhavani.: "Activity Recognition in Egocentric video using SVM, kNN and Combined SVMkNN Classifiers", International Conference on Advanced Material Technologies (ICAMT)-2016.

[17] Xiaodong Yang, Chenyang Zhang, and YingLi Tian.: "Recognizing Actions Using Depth Motion Maps-based Histograms of Oriented Gradients", 20th ACM international conference on Multimedia, October 2012.

[18] Navneet Dalal, Bill Triggs, and Cordelia Schmid.: "Human Detection Using Oriented Histograms of Flow and Appearance", GRAVIR-INRIA, 655 avenue de l'Europe, Montbonnot 38330, France.

[19] Florian Baumann.: "Action Recognition with HOG-OF Features" in: 'Pattern Recognition', pp.243-248.

[20] Cristianini, N. And J. Shawe-Taylor.: "An Introduction to Support Vector Machines and other kernel-based learning methods".