International Journal of Pure and Applied Mathematics
Volume 117 No. 15 2017, 1185-1194
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue
ijpam.eu

Activity Based Human Emotion Recognition in video

R.Santhoshkumar[1]/ Research Scholar
Department of Computer Science and Engineering
Annamalai University
Chidambaram, India
santhoshkumar.ausce@gmail.com

M.Kalaiselvi Geetha[2]/ Associate Professor
Department of Computer Science and Engineering
Annamalai University
Chidambaram, India
geesiv@gmail.com

J.Arunnehru[3]/Assistant Professor
Department of Computer Science and Engineering
SRM University
Chennai, India
arunnehru.ausce@gmail.com

Abstract: In real life situation Human emotion recognition is an attracting great interest for its necessary applications in computer vision area. Human emotions are identified by gesture of body movements. In this paper, an emotion recognition approach based on body movements is proposed. Cumulative Motion Image based Speed Up Robust Feature (CMI-SURF) is extracted as features. The experiments were carried out using emotion dataset considering nine actions (Happy-walking, Happy-sitting, Happy-jumping, Angry-walking, Angry-sitting, Angry-jumping, Fearful-walking, Fearful-sitting and Fearful-jumping) and the unsupervised classifier Hidden Markov Model (HMM) showed the best performance with an overall accuracy rate of 88.2%.

Keywords: Video surveillance; Action recognition; Frame difference; Cumulative motion image; Hidden Markov Model (HMM);

I. INTRODUCTION

Recent research on experimental psychology demonstrated that emotions are important in decision making and rational thinking. Over the years research in emotion recognition mainly concentrated on facial expression, voice analysis, full body movements and gestures. The human body movements expressing fundamental emotions like anger, neutral, happy, fear, disgust, sadness, surprise, etc. Certain body movements are related to specific emotions. For example: fear brings to contrast the body, joy brings to openness and upward acceleration of the fore arms, body turning away is signal of fear and sadness. Body turning towards indicates happiness, anger, surprise. A video surveillance system observes action in open public areas like ATM, banks, railway station, airport, gas stations and commercial buildings for real-time or later analysis which is used to detect crime. Which is aims to recognize and classify the action of a person into certain categories like jumping, sitting, walking, bending and skipping. It is a complex process to recognize human activity due to many factors like speed, shadows, postures, illumination changes and occlusion. The recognition of whole body expressions in significantly durable, because the form of the human body has more degrees of freedom than the face on your own and its overall shape varies strongly during expressed motion.

The large applications of emotion recognition systems in several motivating areas: In surveillance field, to predict and prevent whether a person going to act any suspicious activity, estimate emotional state of students in intelligent tutoring systems, monitor player's motivation and interest in games, social robotics for social interaction, in medical field the applications used for autism and dementia patients by detecting and monitoring depression levels. Apart from these applications, emotion recognition systems find uses in a host of other domains like, Telecommunications, Video Games, Animations, Robotics, and Psychiatry; affect sensitive HCI, Automobile Safety, Educational Software.etc [1] [2].

A. Outline of the Work

This paper deals with human emotion recognition that aims to understand human actions from video sequences and then to identify their emotions while doing these action. The proposed method is evaluated using University of York [12] emotion dataset with the person showing actions such as walking, sitting and jumping with emotion like happy, angry and fearful. Difference image is obtained by subtracting the consecutive frames. Cumulative motion image (CMI) is calculated by combining the four frame difference images and then SURF features are extracted from CMI. The extracted feature is fed to the Hidden Markov Model (HMM). The rest of the paper is organized as follows. Section II explains the related works. Section III explains the workflow of the proposed approach and feature extraction method. Section IV presents the Experimental results. Finally, Section V concludes the paper.

II. RELATED WORKS

The vision based emotion recognition turns into the huge objective to segregate the activities consequently. Recent surveys in the area of human action analysis in [1] and [2] focus on the feature descriptor, representation, and classification model in video sequences. Survey by Turaga et al. [3] centers around recognition of human activity. Wang et al. [4] propose a method which relies on optical flow and edge features, where these two discriminative features were combined to extract the motion and shape descriptors to distinguish one action from another. Reddy et al. [5] present a method to recognize action-based on sphere/rectangle tree structure that is built with spatio temporal interest point features. J. Arunnehru et al. [6] proposed an Automatic human emotion Recognition in surveillance video based on gesture dynamic's features and the features are evaluated by SVM, Naïve Bayes and Dynamic Time Wrapping. Zhu et al. [7] address a multi-view action recognition algorithm based on local similarity random forests and sensor fusion, and normalized silhouettes are used as pose features and effective for multiple view human action recognition. In [8], vocabulary forest is constructed with local static and optical flow features and uses trees and forests for action recognition classification. M. KalaiselviGeetha et al. [9] developed a video retrieval applications for video classification and shot detection using Block intensity comparison code (BICC) and unsupervised shot detection. A noval AANN misclusting Rate (AMR) algorithm is used to detect the shot transitions. Haiyong et al. [10] present a novel classification method to recognize the actions from videos based on centroid-radii model descriptor and to train and classify video sequences by nonlinear SVM decision tree (NSVMDT). J. Arunnehru et al. [11] assigns motion intensity code for action recognition in surveillance video using Region of Interest (ROI) from the difference image. D. Wu, et al. [12] Proposed a method which relies on optical flow and edge features, where these two discriminative features were combined to extract the motion and shape descriptors to distinguish one action from another. J. Arunnehru et al. [13] proposed an application for Action Recognition in automated surveillance. The 18 dimensional Block intensity vector are extracted and evaluated through SVM. Haiyong et al. [14] present a novel classification method to recognize the actions from videos based on centroid-radii model descriptor and to train and classify video sequences by nonlinear SVM decision tree (NSVMDT). In [15], a method was proposed based on extended motion template from human silhouettes. The holistic structural features were extracted from motion templates to discriminate the human action, which represents local and global information. Zhang et al. [10] have used the spatial–temporal with optical flow features. The temporal consistence of motion is improved with an enhanced DTW method to recognize the human actions. Wang et al. H. Bay et al. [22] proposed a novel detector-descriptor scheme, coined SURF (Speeded-Up Robust Features). The detector is based on the Hesian Matrix, but utilizes a simple approximation, for example DoG (Difference of Gaussian) is a simple Laplacian-based detector. Ashwini Ann Varghese, et al [20] proposes an emotion recognition system in real time and describes the advances different types of approaches used for recognizing human emotions. Stefano Piana, et al [21] introduced automatic emotion recognition in real-time from body movements. The real time video are captured and converted into 3D skeletal frames using advance d video capturing system. From the sequences of 3D skeletons, the kinematic, geometrical and postural features are extracted and given to the multi-class SVM classifier to categorise the human emotion.

III. FEATURE EXTRACTION

Feature is a descriptive characteristic extracted from an image or video sequences, which represent the meaningful data that are vital for further analysis. The following subsections present the description of the feature used in this work. The flow diagram of the proposed approach is shown in Fig.1

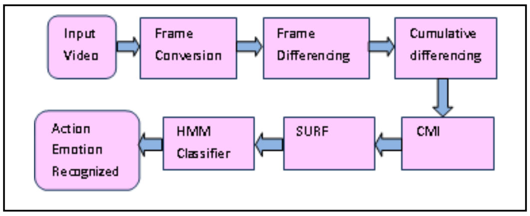

Figure 1: Flow Diagram of the Proposed Approach

A. Frame Differencing

Frame differencing is defined by the difference between two consecutive frames, as an replacement of subtracting a predefined background while in motion, the subtraction of frames model considers every pair of consecutive frames of time t and $t + 1$ to extract any motion details in it. In order to find the regions of interest, by previous frame simply subtracting the current frame on a pixel by pixel model, Fig. 2(a), Fig. 2(b) shows the consecutive frames of the emotion dataset. The frame difference image of the emotion and action is shown in Fig. 2(c). Then the value of the difference image is related with a determine threshold value. The image at time t is given by:

$$I_t(r,s) = |D_t(a,b) - D_{t+1}(a,b)|$$

$$1 < a < w,\ 1 < b < h \qquad (1)$$

$D_t(a, b)$ is the amount of the pixel (a, b) in the emotion dataset frame, w and h are the width and height of the image respectively. The proposed method uses an image size of 960 X 540.

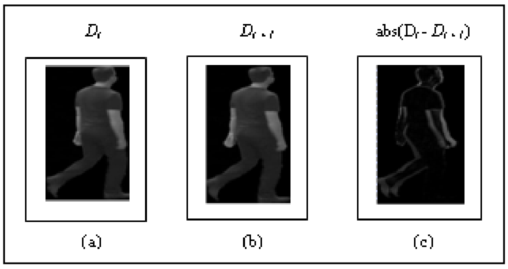

Figure 2: (a) and (b) are two sequence frames. (c) difference frame of (a) and (b)

B. Cumulative differencing

n-frame cumulative differencing is applied to identifying the region as showed in Fig. 3.

$$I_k(r,s) = D_t(r,s) - D_{t+1}(r,s)$$

$$1 \le x \le w,\ 1 \le y \le h \qquad (2)$$

The next step is to calculate then I_k is the difference image found by subtracting by two consecutive frames D_t and D_{t+1}. $D(r,s)$ is the pixel intensity values of (r,s), w and h are width and height of the image respectively. Consecutive difference images are calculated as follows:

$$I_n(r,s) = D_n(r,s) - D_{p+1}(r,s) \qquad (3)$$

$$I_{n+1}(r,s) = D_{p+1}(r,s) - D_{p+2}(r,s) \qquad (4)$$

$$I_{n+2}(r,s) = D_{p+2}(r,s) - D_{p+3}(r,s) \qquad (5)$$

$$I_{n+l}(r,s) = D_{p+l}(r,s) - D_{p+l+1}(r,s) \qquad (6)$$

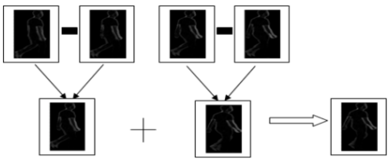

Figure 3: Cumulative differencing frame of walking action

C. Cumulative Motion Images (CMI)

From the video sequence the Motion information is extracted by pixel-wise differencing of successive frames. CMI generation processes for walking action is shown in Fig 4, the moving image is gradually received by the video. CMI is calculated using

$$H(r,s,t) = \begin{cases} \tau & if\ B(r,s,t) = 1 \\ \max(0, H(r,s,t-1) - \delta) & otherwise \end{cases} \quad (7)$$

Where $H(r, s, t)$ is the current CMI, H(r, s, t-1) is the earlier CMI, B(r, s, t) is the current binary image, τ is the maximum value of importance degree, and δ is the reducing value of the importance degree. If the pixel value of the current incoming binary image B(r, s, t) is one, the pixel value of CMI is the maximum value. CMI subtracts the reduction coefficient from the pixel value of the previous CMI; a higher value is then selected after comparing the subtraction value and minimum value (zero). Accordingly, the pixel value where the action is not shown is zero. After the CMI creation, features are executed using SURF.

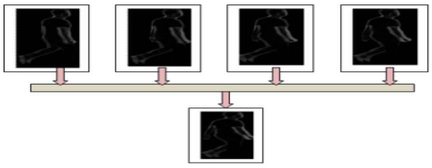

Figure 4: CMI generation process of walking action

D. SURF Feature Extraction

Our technique extricates salient feature and descriptors from images utilizing SURF. Because of its compact descriptor length the SURF extractor is favored over SIFT extractor. However the SIFT execution comprising 128 dimensional feature vector, SURF execution comprising 64 f dimensional feature vector. Quick changes in intensity esteems in both the horizontal and vertical headings are showed by first searching for pixel. Such pixels yield high Harris corner detection scores and are referred to as key points. Key points are searched for over a subspace of {x, y, σ} ϵ R[11]. The key points exists at Gaussian scale space are represented by the variable σ. In SURF, a descriptor vector of length 64 is developed utilizing a histogram of gradient orientation in the neighborhood key points. Figure 4 demonstrates the way in which a SURF descriptor vector is developed. David Lowe furnishes the inclined per user with additional data on neighborhood hearty component extractors [1]. SURF (Speeded-Up Robust Features) is a calculation which makes it conceivable to recognize and depict neighborhood feature of a image. It was exhibited without precedent for [10], however right now is utilized as a part of different frameworks e.g. image description, segmentation[11], image analysis[12], content based image retrieval[13], object tracking, image databases [14], and so on. SURF depends on SIFT, and it utilizes Integral Images rather than DOG (Difference of Gaussian) [12], which enables it to work considerably quicker than SIFT. It can also be quickened by GPU and it has a parallel usage. SURF depends on image key points (fascinating focuses), which makes it conceivable to separate local feature from an image. For each key point, which demonstrates neighborhood image feature, we produce a component vector, which can be utilized for additionally handling. SURF comprises of four principle steps: • Computing Integral Images,

- Fast-Hessian Detector,
  - The Hessian,
  - Constructing the Scale-Space,
  - Accurate Interest Point Localization,
- Interest Point Descriptor,
  - Orientation Assignment,
  - Descriptor Components,
- Generating vectors portraying the key points.

A SURF key points comprises of two vectors [15, 16]. The first contains: point position (x, y), scale (Detected scale), reaction (Response of the distinguished component, quality), Orientation (Orientation measured against clockwise from +ve x-pivot), laplacian (Sign of laplacian for quick coordinating purposes). The second one depicts the intensity appropriation of the pixels inside the area of the intrigue point (64 or 128 esteems). With a specific end goal to produce key points, SURF requires one input parameter minHessian. The technique is impervious to change of scale and pivot, which takes into account coordinating relating key points in comparable images [17]. The most important key points are selected k-means cluster with different k values.

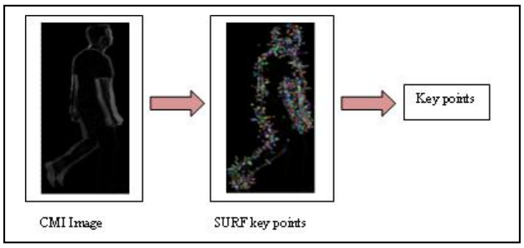

Figure 5: Feature extraction process of CMI-SURF

IV. EXPERIMENTAL RESULTS

Using Emotion dataset (University of York) the proposed method is evaluated in this section. The experiments are carried out in the Opencv 2.4 with Qt Creator in Ubuntu 12.04 LTS operating system on a computer with Intel CORE I7 Processor (3.40GHz), 8.0 GB RAM. The obtained SURF features are fed to Hidden Markov Model (HMM) Classifier to develop a form for each activity and these forms are used to analysis the performances of the classifier.

A. Dataset

Emotion dataset (University of York) is a publicly available dataset, containing four different emotions (happy, angry, fear and sad) performed by 25 actors. The sequences were taken over static (black) background with the frame size of 1920 × 1080 pixels at a rate of 25 fps. For each emotion, actors are performed five different actions: walking, jumping, box picking, box dropping and sitting having an approximate length of 15 seconds of video. In this work, three emotions (happy, angry and fear) and three actions (walking, jumping and sitting) of 10 persons (male and female) are used for experimental purpose. Action recognition with three different emotion are happy, angry, fearful, second row shows the sitting action recognition with three different emotions are happy, angry, fearful, the third row shows the jumping with three different emotion are happy, angry, fearful. For each approximate length of 15 seconds of video obtained 90 data records are considered for experimental purpose. In this work, 10 persons are taken randomly from emotion dataset for evaluation. The samples are divided into a training set of (5 persons), and testing set of (5 persons).

B. Performance Evaluation

The 64-dimensional SURF features are extricated and as illuminated in Sect. 3.1. The Random Forest classifier is assessed utilizing 10-fold cross-validation approach on execution. Precision, Recall (Sensitivity), F-Score and Specificity are the measurements for execution assessment, where, *tp* and *fp* are the quantity of genuine positive and false positive prediction of the class and *tn* and *fn* are the quantity of genuine negative and also negative expectations. Precision gives the general accuracy of the movement recognition. Recall gives an activity is recognized accurately. F-Score is the symphonious mean of Precision and Sensitivity. At last, Specificity provides a measure of strategy is recognizing negative activity accurately. The factual events of Accuracy, Precision, Recall (Sensitivity), F-Score and Specificity is characterized as follows

$$Accuracy = \frac{tp+tn}{tn+fp+tp+fn} \tag{8}$$

$$Precision = \frac{tp}{tp+fp} \tag{9}$$

$$Recall\ (Sensitivity) = \frac{tp}{tp+fn} \tag{10}$$

$$Specificity = \frac{tn}{tn+fp} \tag{11}$$

$$F-Score = 2\frac{Precision\ \ X\ Recall}{Precision\ +Recall} \tag{12}$$

C. Hidden Markov Model

A Hidden Markov Model (HMM) is a statistical model, where the system model is assumed to be a Markov process with unobserved state. It has a finite set of states, each of which is associated with a probability distribution. Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state an outcome or observation can be generated, according to the associated probability distribution. The state is not visible to the external observer and thus the states are „hidden" from outside; hence, it is named as hidden Markov model. It is important to note that the word "hidden" refers to the state sequence through which the model passes, not to the parameters of the model. Even if the model parameters are known exactly, the model is still "hidden".

The hidden Markov model consists of two stage stochastic processes: One is an unobservable Markov chain with a finite number of states, an initial state probability distribution and a state transition probability matrix, and the other is a set of probability density functions associated with each state. The probability density function can be either discrete (discrete HMM) or continuous (continuous HMM). The continuous HMM is characterized by the following [13]:
• Mn, the number of states in the model. The individual states are denoted as
s = s1, s2, sMn, and the state at time t as qt.
• The state transition probability distribution A = brs , where

$$a_{rs} = X[Yt + 1 = js \mid Yt = jr], 1 \le r, s \le Mn$$

Defines the Probability transition from state js to jr at time t

$$\sum_{s=1}^{nx} a_{rs} = 1, 1 \le r \le nx$$

The observation probability density function in state r,

$$h_r(O) = \sum_{K=1}^{N} D_{rk}, 1 \le r \le Mn$$

where D_{rk} is the mixture coefficient for k^{th} mixture component in state *r*. *N* is the number of components in a Gaussian mixture model, and $E(O, \mu_{rk}, \sum_{rk})$ is a Gaussian probability density function with mean μ_{rk} and covariance $\sum_{rk}$
The initial state distribution $\prod = \pi$, where

$$\prod_r = E[F_1 = js], \qquad 1 \le r \le Mn$$

HMM assumes that an input observation sequence of feature vector follows a multi-state distribution. It is expressed by the initial state distribution (π), state transition probabilities (A), and observation probability distribution in each state (B). In HMM training, it estimates the parameter set denoted by a compact notation, $\alpha = (A, B, \pi)$for each class based on the training sequences. It enters a new state based on the transition probability depending on the previous state. After making the transition depending on the current state an output symbol is produced based on the probability distribution.

Hidden Markov model (HMM) is a doubly embedded stochastic process where the underlying stochastic process is not directly observable [13]. It is a statistical model in which the system modeled is assumed to be a Markov process with unobserved state, especially to model temporal features. The HMM is not only a model underlying feature vector but also the temporal sequencing of the features. In a regular Markov model, the state is directly visible to the observer, and its parameters are to state transition probabilities. In a hidden Markov model, the state is not directly visible. Each state has a probability distribution over the possible output tokens. Therefore the sequence generated by a HMM gives some details about the sequence of states. In this work, three, four and five states are considered. Experiments are conducted on Emotion dataset considering nine different actions viz (Happy-walking, Happy-sitting, Happy-jumping, Angry-walking, Angry-sitting, Angry-jumping, Fearful-walking, Fearful-sitting and Fearful-jumping).

The recognition results obtained by the proposed method on emotion dataset with HMM is summarized in a confusion matrix in Table 1 and Table 2, where correct responses define the main diagonal, the majority of actions are correctly classified, An average recognition rate of HMM is **88.2%** for k=3 and 86.1% for k=5 in 3 state HMM. Table 3 shows the performance measure obtained with *k* means cluster k=3 and k=5, of 3 state HMM.

Table I: Confusion Matrix for 3 state HMM (*k* means cluster, k=3)

Class	*Happy (Walk)*	*Happy (Sit)*	*Happy (Jump)*	*Angry (Walk)*	*Angry (Sit)*	*Angry (Jump)*	*Fearful (Walk)*	*Fearful (Sit)*	*Fearful (Jump)*
Happy (Walk)	**5**	0	0	0	0	0	0	0	0
Happy (Sit)	0	**5**	0	0	0	0	0	0	0
Happy (Jump)	0	0	**5**	0	0	0	0	0	0
Angry (Walk)	**1**	0	0	**4**	0	0	0	0	0
Angry (Sit)	0	**1**	0	0	**3**	0	0	**1**	0
Angry (Jump)	0	0	**1**	0	0	**3**	0	0	**1**
Fearful (Walk)	0	0	0	0	0	0	**5**	0	0
Fearful (Sit)	0	0	0	0	0	0	0	**5**	0
Fearful (Jump)	0	0	0	0	0	0	0	0	**5**

Table II: Confusion Matrix for 3 state HMM (*k* means cluster, k=5)

Class	*Happy (Walk)*	*Happy (Sit)*	*Happy (Jump)*	*Angry (Walk)*	*Angry (Sit)*	*Angry (Jump)*	*Fearful (Walk)*	*Fearful (Sit)*	*Fearful (Jump)*
Happy (Walk)	**5**	0	0	0	0	0	0	0	0
Happy (Sit)	0	**5**	0	0	0	0	0	0	0
Happy (Jump)	0	0	**5**	0	0	0	0	0	0
Angry (Walk)	**1**	0	0	**4**	0	0	0	0	0
Angry (Sit)	0	**1**	0	0	**3**	0	0	0	**1**
Angry (Jump)	0	0	0	**1**	0	**3**	**1**	0	0
Fearful (Walk)	0	0	0	0	0	0	**5**	0	0
Fearful (Sit)	0	0	0	0	0	0	0	**5**	0
Fearful (Jump)	0	0	0	0	0	0	0	0	**5**

Table III: Performance measure observed with *k*-means cluster k=3 and k=5 of 3 state HMM

Emotion/ Action	*Precision (%)*		*Recall (%)*		*Specificity(%)*		*F-Measure (%)*	
	k=3	*k=5*	*k=3*	*k=5*	*k=3*	*k=5*	*k=3*	*k=5*
Happy (Walk)	83.3	83.5	100	98.5	97.5	96.4	90.9	88.7
Happy (Sit)	83.3	83.5	100	98.5	97.5	96.4	90.9	88.7
Happy (Jump)	83.3	99	100	98.5	97.5	100	90.9	89.1
Angry (Walk)	100	88	80	99.2	100	100	88.9	86.3
Angry (Sit)	100	99	60	99.2	100	100	75	73.2
Angry (Jump)	100	99	60	99.2	100	96.6	75	73.2
Fearful (Walk)	100	83.5	100	75.6	100	96.6	100	99
Fearful (Sit)	83.3	83.5	100	75.6	97.5	100	90.9	89.1
Fearful (Jump)	83.3	98	100	100	97.5	96.4	90.9	89.1
Average	**90.7**	**91.1**	**88.9**	**93.8**	**98.6**	**97.1**	**88.2**	**86.1**

Table IV: Confusion Matrix for 4 state HMM (*k* means cluster, k=3)

Class	*Happy (Walk)*	*Happy (Sit)*	*Happy (Jump)*	*Angry (Walk)*	*Angry (Sit)*	*Angry (Jump)*	*Fearful (Walk)*	*Fearful (Sit)*	*Fearful (Jump)*
Happy (Walk)	**5**	0	0	0	0	0	0	0	0
Happy (Sit)	0	**4**	0	0	0	0	0	**1**	0
Happy (Jump)	**1**	0	**3**	0	0	0	0	0	0
Angry (Walk)	**4**	0	0	**1**	0	0	0	0	0
Angry (Sit)	0	**2**	0	0	**2**	0	0	**1**	0
Angry (Jump)	0	0	**1**	0	0	**2**	0	0	**2**
Fearful (Walk)	0	0	**1**	0	0	0	**4**	0	0
Fearful (Sit)	0	0	0	0	0	0	0	**4**	0
Fearful (Jump)	0	0	0	0	0	0	0	0	**5**

Table V: Confusion Matrix for 4 state HMM (*k* means cluster, k=5)

Class	*Happy (Walk)*	*Happy (Sit)*	*Happy (Jump)*	*Angry (Walk)*	*Angry (Sit)*	*Angry (Jump)*	*Fearful (Walk)*	*Fearful (Sit)*	*Fearful (Jump)*
Happy (Walk)	**5**	0	0	0	0	0	0	0	0
Happy (Sit)	0	**3**	0	0	**1**	0	0	**1**	0
Happy (Jump)	**1**	0	**3**	0	0	0	0	0	**1**
Angry (Walk)	**3**	0	0	**1**	0	0	**1**	0	0
Angry (Sit)	0	**2**	0	0	**2**	0	0	**1**	0
Angry (Jump)	0	0	**1**	0	0	**2**	0	0	**2**
Fearful (Walk)	0	0	**1**	**1**	0	0	**3**	0	0
Fearful (Sit)	0	0	0	0	**1**	0	0	**4**	0
Fearful (Jump)	0	0	0	0	0	0	0	0	**5**

The emotion and action recognition results obtained by the proposed method on emotion dataset with HMM is summarized in a confusion matrix in Table 4 and Table 5, where correct responses define the main diagonal, the majority of actions are correctly classified, An average recognition rate of HMM is 64% for k=3 and 61.4% for k=5 in 4 state HMM. Table 6 shows the performance measure obtained with *k* means cluster, k=3 and k=5 of 4 state HMM. From this measurement *k* means cluster, k=3 with 3 state given a better performance than others.

Table VI: Performance measure observed with *k*-means cluster k=3 and k=5 of 4 state HMM

Emotion/ Action	*Precision (%)*		*Recall (%)*		*Specificity(%)*		*F-Measure (%)*	
	k=3	*k=5*	*k=3*	*k=5*	*k=3*	*k=5*	*k=3*	*k=5*
Happy (Walk)	50	55.2	100	95.1	87.5	100	66.7	65
Happy (Sit)	66.7	62.8	80	91.6	95	90	72.7	65
Happy (Jump)	60	62.1	60	91.6	95	90	60	52.4
Angry (Walk)	100	89.2	80	91.2	100	90	33.3	58.2
Angry (Sit)	66.7	64.5	40	45.1	97.5	87.6	50	30
Angry (Jump)	100	88	40	50	100	79.5	57.1	30
Fearful (Walk)	80	91.1	80	61.5	97.5	95.6	80	75.6
Fearful (Sit)	66.7	65	80	61.5	95	95.6	72.7	94.2
Fearful (Jump)	71.4	70	100	90.7	95	90	83.3	82.2
Average	**73.5**	**71.9**	**66.7**	**75.3**	**95.8**	**90.5**	**64**	**61.4**

V. CONCLUSION AND FUTURE WORK

This paper proposed an effective structure for emotion recognition task using CMI-SURF (Speeded Up Robust Feature). The experiments are evaluated on Emotion dataset. The execution of CMI-SURF feature in video processing is assessed utilizing Hidden Markov Model (HMM) classifier. From the test, the Hidden Markov Model demonstrates a recognition exactness of 88.2% for Emotion dataset. The exploratory outcomes have exhibited that the CMI-SURF strategy has a promising execution on Emotion dataset. In this experiment, the activities which is confused and difficult to separate is take it as future challenge and it needs promote consideration.

References

[1] R. Poppe, Vision-based human motion analysis: an overview. Comput. Vis. Image Underst. (CVIU) 108(1–2), 4–18 (2007).

[2] R. Poppe, A survey on vision-based human action recognition. IVC 28, 976–990 (2010).

[3] P. Turaga, R. Chellappa, S. Venkatramana Subrahmanian, O. Octavia Udrea, Machine recognition of human activities: a survey. IEEE Trans. Circ. Syst. Video Technol. 18(11), 1473–1488 (2008).

[4] L. Wang, Y. Wang, T. Jiang, D. Zhao, W. Gao, Learning discriminative features for fast frame-based action recognition. Pattern Recogn. 46(7), 1832–1840 (2013).

[5] K. Reddy, J. Liu, M. Shah, Incremental action recognition using feature-tree, in International Conference on Computer Vision (2009).

[6] J. Arunnehru and M. Kalaiselvi Geetha. Automatic Activity Recognition for Video Surveillance. International Journal of Computer Applications, Published by Foundation of Computer Science, New York, USA, 75(9):1-6, August 2013.

[7] F. Zhu, Ling Shao, Mingxiu Lin.: Multi-view action recognition using local similarity random forests and sensor fusion. Pattern Recogn.Lett. 34(1), 20–24 (2013).

[8] Z. Lin, Z. Jian, L. Davis, Recognizing actions by shape motion prototype trees, in International Conference on Computer Vision (2009).

[9] M. Kalaiselvi Geetha, S. Palanivel, Video Classification and Shot Detection for Video Retrieval Applications, International Journal of Computational Intelligence Systems, vol. 2, no. 1, pp. 39-50, March 2009.

[10] H. Zhao, Z. Liu, Human action recognition based on non-linear SVM decision tree. J. Computat. Infor. Syst. 7, 2461– 2468 (2011).

[11] J. Arunnehru, M. Kalaiselvi Geetha, Motion Intensity Code for Action Recognition in Video Using PCA and SVM, Lecturer Notes in Artificial Intelligence (LNAI), Springer International Publishing, Switzerland, Vol. 8284, pp. 70–81, December 2013.

[12] D. Wu, L. Shao, Silhouette analysis-based action recognition via exploiting human poses. IEEE Trans. Circuits Syst. Video Techn. 23(2), 236–243 (2013).

[13] J. Arunnehru, and M. Kalaiselvi Geetha, Automatic Human Emotion Recognition in Surveillance Video. Intelligent Techniques in Signal Processing for Multimedia Security, Springer-Verlag, pp. 321-342, Oct. 2017.

[14] W. Zhang, Y. Zhang, C. Gao, J. Zhou, Action recognition by joint spatial-temporal motion feature. J. Appl. Math. (2013).

[15] L. Wang, Y. Wang, T. Jiang, D. Zhao, W. Gao, Learning discriminative features for fast frame based action recognition. Pattern Recogn. 46(7), 1832–1840 (2013).

[16] M. Blank, L. Gorelick, E. Shechtman, M. Irani, R. Basri, Actions as space-time shapes, in Proceedings of IEEE International Conferences on Computer Vision (2005), pp. 1395–1402.

[17] Eum, Hyukmin, et al. “Continuous Human Action Recognition Using Depth-MHI-HOG and a Spotter Model.” Sensors 15.3 (2015): 5197-5227.

[18] J.R. Quinlan, C4.5: Programs for Machine Learning (Morgan Kaufmann Publishers, Burlington, 1993).

[19] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (Morgan Kaufmann Publishers, Burlington, 1999).

[20] Ashwini Ann Varghese, et al. Overview on Emotion Recognition System, 2015 International Conference on Soft-Computing and Network Security (ICSNS -2015), Feb. 25 – 27, 2015, Coimbatore, INDIA.

[21] Stefano Piana, et al. “Real-time automatic emotion recognition from body gestures” Human-Computer Interaction (cs.HC); Computer Vision and Pattern Recognition (cs.CV), 2014.

[22] H. Bay, T. Tuytelaars, and L. Van Gool “SURF: Speeded Up Robust Features”, IEEE Trans. Pattern Anal. Mach. Intell. 23(3), 257–267, 2007.

Rajesh, M. & Gnanasekar, J.M. Wireless Pers Commun (2017),https://doi.org/10.1007/s11277-017-4565-9