International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP. 88-97

Recognition of Emotions from Human Activity Using STIP Feature

*R. Santhoshkumar[1], M. Kalaiselvi Geetha[2]
Research Scholar, Associate Professor
Department of Computer Science and Engineering, Annamalai University Chidambaram, India

Abstract: *Recently, increasing attention has been paid to the detection of spatio-temporal interest points (STIPs), which has become a key technique and research focus in the field of computer vision. The Local image features or interest points provide compact and abstract representations of patterns in frames. In this paper, an emotion recognition approach based on body movements is proposed. The videos are converted into frames and the frames are preprocessed using median filter. The Harris method is used for corner detection in human present in the frame. Then the STIP features are extracted for the human present in the frames. Similarly the trained and test data are extracted using this STIP feature. The extracted features are fed to the KNN and Support Vector Machine (SVM) classifiers to classify the emotion of the human. The performance measure can be calculated using F-Score. The Emotion dataset are used for this experiment.*

Keywords: *Spatio-temporal interest points (STIPs), Harris, Support Vector Machine (SVM) classifier, Emotion Recognition*

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, and Educational Software.etc [1] [2].

1.1Outline 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 five frame difference images and then SURF features are extracted from CMI. The extracted feature is fed to the Support Vector Machine (SVM) and k Nearest Neighbor (KNN). The rest of the paper is organized as follows. Section 2 explains the related works of the paper. Section 3 explains the workflow of the proposed approach and feature extraction method. Section 4 presents the Experimental results. Finally, Section 5 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. Kalaiselvi Geetha 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. [16] 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. [17] 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. R.Santhoshkumar, M.Kalaiselvi Geetha, et al [23] introduces a robust HoG (Histogram of Oriented Gradients) feature and KLT tracker for emotion recognition in video. The main contribution of this work is to build HoG (Histogram of Oriented Gradients) features with different bins for emotion recognition from human in the video. R.Santhoshkumar, M.Kalaiselvi Geetha, et al [24] an emotion recognition approach based on body movements is proposed. Human emotions are identified by gesture of body movements. Cumulative Motion Image based Speed up Robust Feature (CMI-SURF) is extracted as features.

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

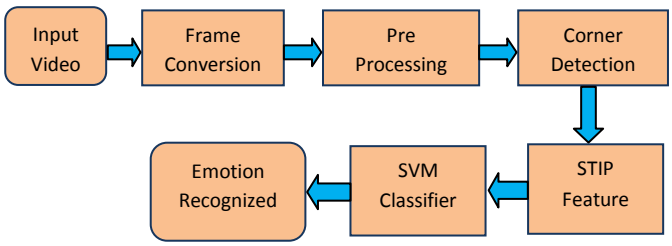

Fig. 1: Flow Diagram of the Proposed Approach

2. 1Load Video and Frame conversion

From the emotion dataset the videos are loaded one by one and converted into frames. As shown in fig.

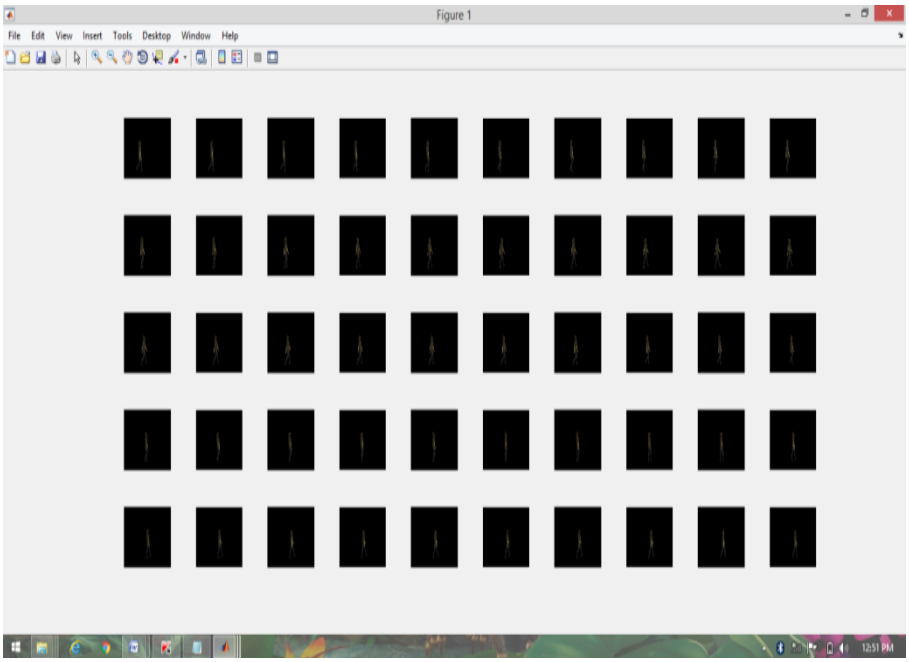

Fig. 2: Video into frames of happy walking

2.2Preprocessing

The sequences of frames are preprocessed using median filter for removing noise in the frames. The figure 3 had shown the preprocessed image.

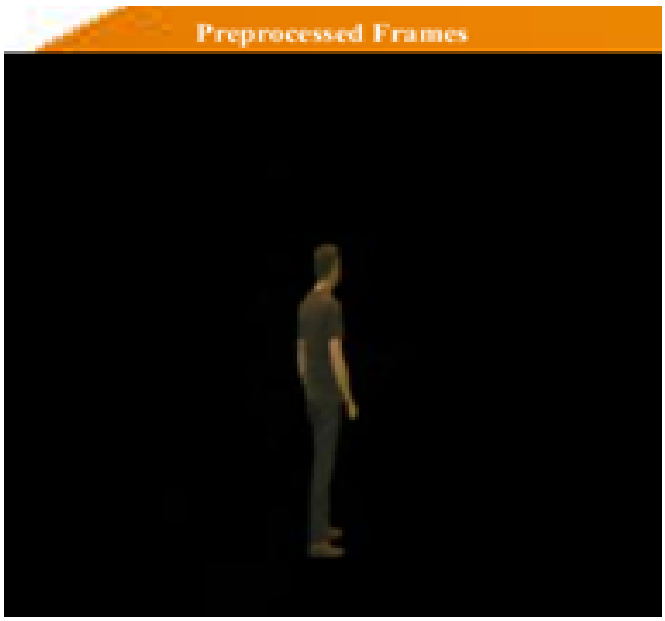

Fig. 3: Preprocessed frames

A. Spatio-Temporal Interest Point Extraction

Interest Points in the Spatial DomainIn the spatial domain, we can model an image $f^{sp}: R^2 \rightarrow R$ by its linear scale-space representation (Witkin, 1983; Koenderink and van Doorn, 1992; Lindeberg, 1994; Florack, 1997) M^{sp}:$R^2 \times R+ \rightarrow R$

$$M^{sp}(a,b;\sigma_m^2) = h^{sp}(a,b;\sigma_m^2) * f^{sp}(a,b) \quad (1)$$

defined by the convolution of f^{sp} with Gaussian kernels of variance σ_m^2

$$h^{sp}(a,b;\sigma_m^2) = \frac{1}{2\pi\sigma_m^2} exp(-(a^2+b^2)/2\sigma_m^2) \quad (2)$$

The idea of the Harris interest point detector is to find spatial locations where *f sp* has significant changes in both directions. For a given scale of observation σ_m^2 , such points can be found using a second moment matrix integrated over a Gaussian window with variance σ_i^2 (Forstner and Gulch, 1987; Bigun et al., 1991; Garding and Lindeberg, 1996):

$$\mu^{sp}(.;\sigma_m^2,\sigma_i^2) = h^{sp}(.;\sigma_i^2) * \left(\left(\nabla M(.;\sigma_m^2)\right)\left(\nabla M(.;\sigma_m^2)\right)^T\right) \quad (3)$$

$$= h^{sp}(.;\sigma_i^2) * \begin{pmatrix} \left(M_a^{sp}\right)^2 & M_a^{sp}M_b^{sp} \\ M_a^{sp}M_b^{sp} & \left(M_b^{sp}\right)^2 \end{pmatrix} \quad (4)$$

where '$*$ ' denotes the convolution operator, and M_a^{sp} and M_b^{sp} are Gaussian derivatives computed at local scale σ_m^2 according to $M_a^{sp} = \partial a\ (h^{sp}\ (\cdot;\sigma_m^2) * f^{sp}\ (\cdot))$ and $M_b^{sp} = \partial b\ (h^{sp}\ (\cdot;\sigma_m^2\ ) * f^{sp}\ (\cdot))$. The second moment descriptor an be thought of as the covariance matrix of a two-dimensional distribution of image orientations in the local neighborhood of a point. Hence, the eigen values λ_1, λ_2 of μ^{sp} constitute descriptors of variations in f^{sp} along the two image directions. Specifically, two significantly large values of λ_1, λ_2 indicate the presence of an interest point. To detect such points, Harris and Stephens (1988) proposed to detect positive maxima of the corner function

$$I^{sp} = \det(\mu^{sp}) - K\ trace^2(\mu^{sp}) \quad (5)$$

$$= \lambda_1\lambda_2 - K(\lambda_1+\lambda_2)^2 \quad (6)$$

At the positions of the interest points, the ratio of the eigen values $\alpha = \lambda_2/\lambda_1$ has to be high. From (4) it follows that for positive local maxima of H^{sp}, the ratio α has to satisfy $k \leq \alpha/(1+\alpha)2$. Hence, if we set $k = 0.25$, the positive maxima of H will only correspond to ideally isotropic interest points with $\alpha = 1$, i.e. $\lambda_1 = \lambda_2$. Lower values of k allow us to detect interest points with more elongated shape, corresponding to higher values of α. A commonly used value of k in the literature is $k = 0.04$ corresponding to the detection of points with $\alpha <$ 23. The result of detecting Harris interest points in image sequence of an anger person is shown in Fig. 2.

1. Interest Points in the Spatio-Temporal Domain

In this section, we develop an operator that responds to events in temporal image sequences at specific locations and with specific extents in space-time. The idea is to extend the notion of interest points in the spatial domain by requiring the image values in local spatio-temporal volumes to have large variations along both the spatial and the temporal directions. Points with such properties will correspond to spatial interest points with distinct locations in time corresponding to local spatio-temporal neighborhoods with non-constant motion. To model a spatio-temporal image sequence, we use a function f : R2 × R → R and construct its linear scale-space representation M:R2 × R × R2 + → R by

$$M(.;\sigma_m^2,\tau_m^2) = h(.;\sigma_m^2,\tau_m^2) * f(.) \quad (7)$$

convolution of f with an anisotropic Gaussian kernel1 with independent spatial variance σ_m^2 and temporal variance τ where the spatio-temporal separable Gaussian kernel is defined as

$$h(a,b,t;\ \sigma_m^2,\tau_m^2) = \frac{1}{\sqrt{(2\pi)^3\sigma_m^4\tau_m^2}} \quad (8)$$

$$X \exp(-(a^2+b^2)/2\sigma_m^2 - t^2/2\tau_m^2\ ) \quad (9)$$

Using a separate scale parameter for the temporal domain is essential, since the spatial and the temporal extents of events are in general independent. Moreover, events detected using our interest point operator depend on both the spatial and the temporal scales of observation and, hence, require separate treatment of the corresponding scale parameters σ_m^2 and τ_m^2 . Similar to the spatial domain, we consider a spatiotemporal second-moment matrix, which is a 3-by-3 matrix composed of first order spatial and temporal derivatives averaged using a Gaussian weighting function

$$h(.;\sigma_i^2\tau_i^2) \quad (10)$$

$$\mu = h(.;\sigma_i^2,\tau_i^2) * \begin{pmatrix} M_a^2 & M_aM_b & M_aM_t \\ M_aM_b & M_b^2 & M_bM_t \\ M_aM_t & M_aM_t & M_t^2 \end{pmatrix} \quad (11)$$

where we here relate the integration scales σ_i^2 and τ_i^2 to the local scales σ_m^2 and τ_m^2 according to $\sigma_i^2 = s\sigma_m^2$ and $\tau_i^2 = s\tau_m^2$. The first-order derivatives are defined as

$$M_a(\sigma_m^2,\tau_m^2) = \partial_a(h * f) \quad (12)$$

$M_b(\sigma_m^2, \tau_m^2) = \partial_b(h * f)$ (13)

$M_t(\sigma_m^2, \tau_m^2) = t(h * f)$ (14)

To detect interest points, we search for regions in f having significant eigenvalues $\lambda 1$, $\lambda 2$, $\lambda 3$ of μ. Among different approaches to find such regions, we propose here to extend the Harris corner function (4) defined for the spatial domain into the spatio-temporal domain by combining the determinant and the trace of μ as follows:

$H = det(\mu) = K\ trace^3(\mu)$ (15)

$\lambda_1\lambda_2\lambda_3 = k(\lambda_1+\lambda_2+\lambda_3)$ (16)

To show how positive local maxima of H correspond to points with high values of $\lambda 1$, $\lambda 2$, $\lambda 3$ ($\lambda 1 \leq \lambda 2 \leq \lambda 3$), we define the ratios $\alpha = \lambda 2/\lambda 1$ and $\beta = \lambda 3/\lambda 1$ and re-write H as

$H = \lambda_1^3(\alpha\beta - k(1+\alpha+\beta)^3)$ (17)

From the requirement $H \geq 0$, we get $k \leq \alpha\beta/(1+\alpha+\beta)3$ and it follows that k assumes its maximum possible value $k = 1/27$ when $\alpha = \beta = 1$. For sufficiently large values of k, positive local maxima of H correspond to points with high variation of the image values along both the spatial and the temporal directions. In particular, if we set the maximum value of α, β to 23 as in the spatial domain, the value of k to be used in H (8) will then be $k \approx 0.005$. Thus, spatio-temporal interest points of f can be found by detecting local positive spatio-temporal maxima in H. Using this method the STIP feature are extracted, which is shown in the fig. 2.

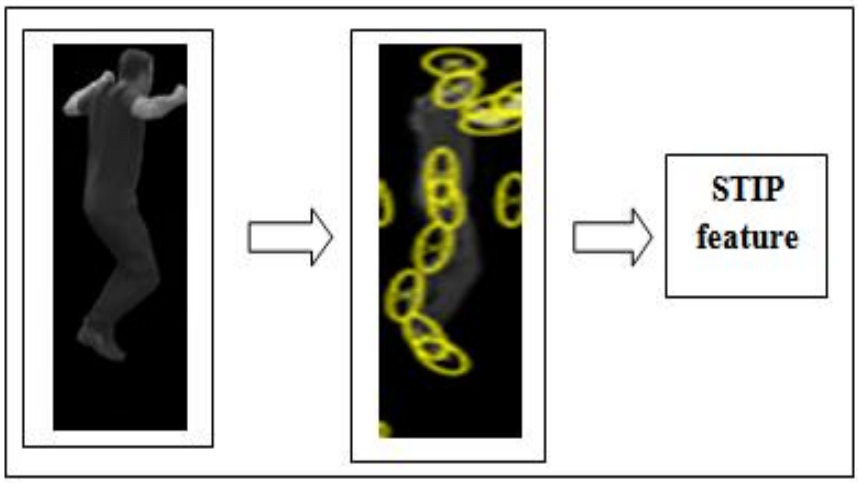

Fig. 4: Extraction of STIP feature for angry jumping from gray frame

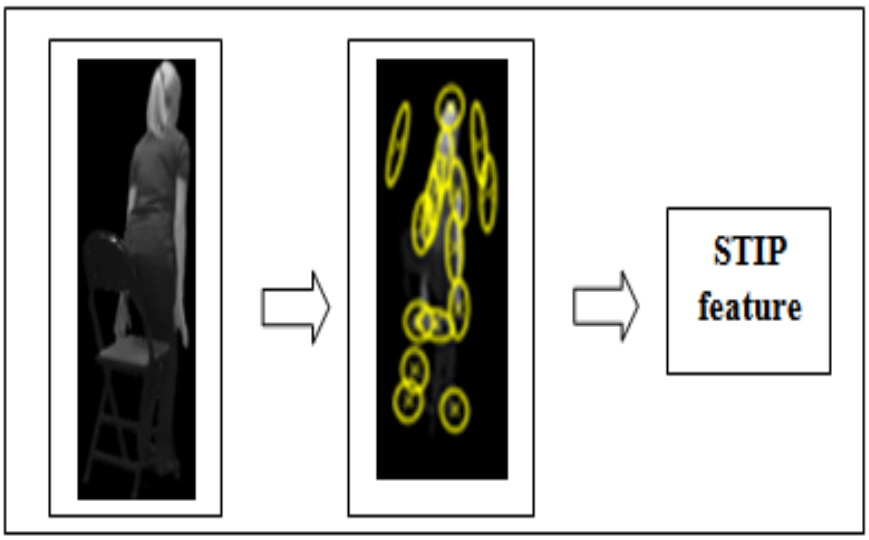

Fig. 5: Extraction of STIP feature for fear sitting from gray frame

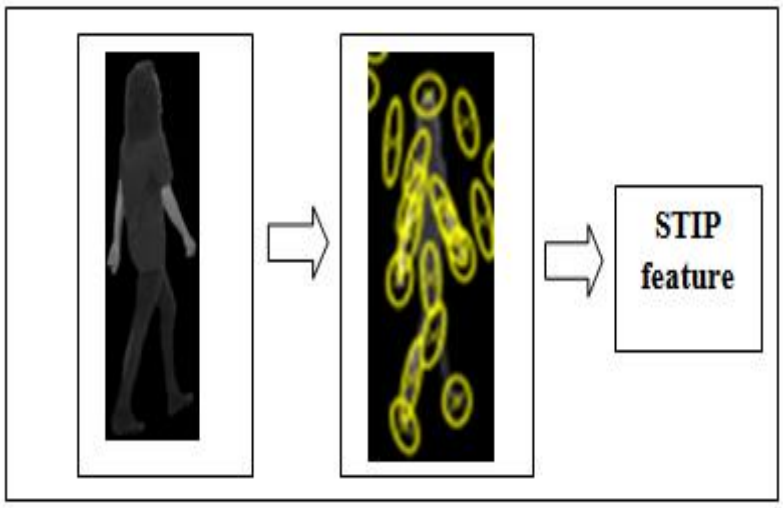

Fig. 6: Extraction of STIP feature for happy walking from gray frame

IV. Experimental Results

The successive five emotions and three actions are mind for emotion recognition: angry, fear, happy, sad and untrustworthy from Emotion dataset. The demonstrations are conducted on Windows 7 Operating System using MATLAB 2015a on a computer with Intel Core i7 Processor 3.40 GHz with 8 GB RAM. The Support Vector Machines (SVM) and k Nearest Neighbor classifiers are used for recognizing the action emotion from Emotion dataset.

V. Emotion 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, five emotions (angry, fear, happy, sad and untrustworthy) and three actions (jumping, sitting, and walking) of 10 persons (male and female) are used for experimental purpose. Action recognition with five different emotion are angry, fearful, happy, sad and untrustworthy, second row shows the sitting action recognition with three different emotions are angry, fear, happy, sad and untrustworthy, the third row shows the jumping with three different emotion are angry, fear, happy, sad and untrustworthy. 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 (6 persons), and testing set of (4 persons).

5.1 Performance Evaluation

The STIP features were extracted for all action emotion in the dataset. The training and testing set are given to the KNN and SVM classifiers one by one. Accuracy, Recall, F-Score, Specificity and Precision are the measuring assessment for this execution. Accuracy is a measure of precision. Recall provides how extraordinary an emotion is recognized accurately. The symphonious mean of Precision and Recall is called as F-score. Specificity shows an evaluation of how great a strategy is recognizing negative emotion accurately. At last, Precision shows the general accuracy of the movement recognition. The factual evaluation of Accuracy, Recall, F-Score, Specificity and Precision are given as follows:

$$Accuracy = \frac{tp+tn}{tn+fp+tp+fn} \quad (18)$$

$$Recall = \frac{tp}{tp+fn} \quad (19)$$

$$F-Score = 2\frac{Precision\ X\ Recall}{Precision\ +Recall} \quad (20)$$

$$Specificity = \frac{tn}{tn+fp} \quad (21)$$

$$Precision = \frac{tp}{tp+fp} \quad (22)$$

where, *tp* and *tn* are the quantity of true positive and true negative prediction of the class and *fp* and *fn* are the quantity of false positive and false negative expectations.

5.3 Support Vector Machine (SVM)

The Support Vector Machine (SVM) is an important and efficient technique for classification in visual pattern recognition [19], [20]. The SVM is most extensively used in kernel learning algorithm. The elegant theory used to separate two classes by large-margin hyper planes. It cannot be extended easily to separate N mutually exclusive classes. The most popular "one-vs-others" approach is used for the multi class problem where, one class is separated from N classes. The classification task are typically involves with training and testing data. The training data are separated by (s_1, t_1), (s_2, t_2), ...(s_n, t_n) into two classes, where $b_j \in \{+1,-1\}$ are the class labels and $s_j \in t_N$ contains n-dimensional feature vector. The goal of Support Vector Machine is to develop a model which predicts target value from testing set. $w.s + b = 0$ is the hyper plane of binary classification, where $w \in R^N$, The two classes are separated by $b \in R$ [23].

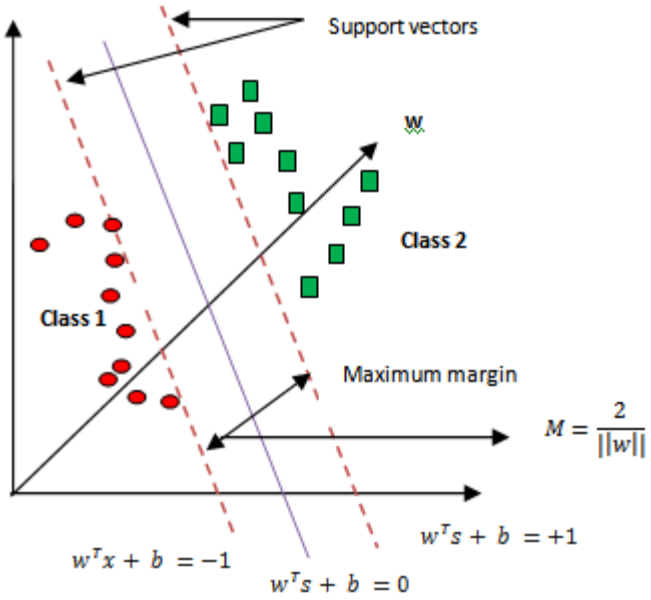

Fig. 7: Illustration of hyperplane in linear SVM

M = 2/||*w*|| is the large margin as show in Fig. 3.The Lagrange multipliers α_i (i=1, ...m) are used to solve the minimization problem, where *v* and *y* are optimal values obtained from Eq. 12.

$$h(s) = sgn\left(\sum_{j=1}^{n} x_j b_j L(s_j, s) + y\right) \quad (23)$$

Maximize the margin and minimize the training error using non-negative slack variables ε_j . The Eq. 13 and Eq. 14 obtain the soft margin Classifier.

$$\min_{v,y,\varepsilon} \frac{1}{2} v^R v + D \sum_{j=1}^{k} \varepsilon_j \quad (24)$$

$$b_j (v^R \emptyset(s_j) + y \geq 1 - \varepsilon_j, \varepsilon_j \geq 0 \quad (25)$$

When the training sample is not linearly separable, the input space mapped into high dimensional space using kernel function $L(s_j s_k) = \emptyset(s_j).\ \emptyset(s_k)$ [20].

Linear: $$L(s_j s_k) = s^R j^s k \quad (26)$$

Polynomial: $$L(s_j, s_k) = (\alpha s^R j^s k + \alpha)^c, \alpha > 0 \quad (27)$$

Radial Basis Function (RBF):

$$L(s_j, s_k) = \exp(-\alpha ||s_j - s_k||^2), \alpha > 0 \quad (28)$$

Sigmoid: $$L(s_j, s_k) = \tanh(\alpha s^R j - s_k + t) \quad (29)$$

Where, α, t, and c are parameters of kernel.

The multiclass Support Vector Machine (SVM) is constructed by N-binary classifiers and one class was separated from rest of the class. Here "one-vs-others" approach is used in this SVM. The five classes of emotions are used in this work. The j^{th} class of the training sets have positive labels and all others with negative labels. The j^{th} SVM solves j^{th} decision function given in Eq. 23. Finally, the feature vectors from the STIP feature are given into multiclass-SVM for classification of human emotion.

5.4 K Nearest Neighbor

K nearest neighbor algorithm have been used since 1970 in many fields like statistical estimation and pattern recognition etc. *K* nearest neighbor is a simple and popular technique for pattern recognition field. It is a type of supervised learning method. It is said to be a lazy learning where the function is only approximated locally. *K* nearest neighbor is a non-parameter algorithm where samples are classified depends on the category of their nearest neighbor. According to the *k* learning samples, the classification algorithm finds the test sample's categories which are the nearest neighbor to the test sample. However, the classification algorithm needs to compute all distance between training sample and testing sample. The process of *K* nearest neighbor algorithm to categorize sample S. Assume *q* training samples A1,A2,…,A*q*. After feature reduction N is the addition of the training samples and get n-dimension feature vector. The all training samples (S1, S2,…, Sn) have the same feature vector of sample S and evaluate the similarities among them. For example taking the p^{th} sample b_p (b_p1, b_p2,…,b_{pn}) and the similarity SIM(S, b_p) is:

$$SIM(S, b_p) = \frac{\sum_{q=1}^{n} S_q . b_{pq}}{\sqrt{\left(\sum_{q=1}^{n} S_q\right)^2} . \sqrt{\left(\sum_{q=1}^{n} b_{pq}\right)^2}} \quad (30)$$

The Larger N similarities, SIM(S, bp), (p=1, 2,…, N), of *k* samples are chosen and consider them as a *K* nearest neighbor collection of S. Then, the probability of S can be calculated using this formula.

$$R(S, A_q) = \sum_b (S, b_p) . y(b_p, A_q) \quad (31)$$

Where y(b_p, A_q) is a category attribute function.

5.5 Evaluation of Emotion Dataset

The normal recognition exactness is **92.4 %** on the Emotion dataset and the confusion matrix appeared in Fig.4. The corner to corner of the confusion matrix illustrates the percentage of instance that was classified accurately. The each emotion class occurrence is spoken to by the lines and the emotion class anticipated by the classifier is spoken to by the sections. The emotions like sad, fear and pride are grouped well with precision more noteworthy than 90%. From this, angry and joy emotions are confused as curve, where these two emotions instinctively appear to be difficult to separate and it needs promote consideration. The execution assessment comes about are computed for the evaluated STIP feature has a better exactness, recall, F-score and Specificity for KNN and SVM with RBF kernel classifiers on Emotion dataset.

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

Class	AJ	AS	AW	FJ	FS	FW	HJ	HS	HW	SJ	SS	SW	UJ	US	UW
Angry (Jump)	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Angry (Sit)	0	38	0	0	0	0	0	1	0	0	1	0	0	0	0
Angry (Walk)	0	0	38	0	1	0	0	0	1	0	0	0	0	0	0
Fearful (Jump)	0	0	0	40	0	0	0	0	0	0	0	0	0	0	0
Fearful (Sit)	1	0	1	1	35	0	0	1	0	1	0	0	0	0	0
Fearful (Walk)	0	0	0	0	0	39	0	0	0	0	0	0	0	0	1
Happy (Jump)	0	0	0	0	0	0	39	0	0	0	0	1	0	0	0
Happy (Sit)	0	0	0	0	0	0	0	40	0	0	0	0	0	0	0
Happy (Walk)	0	0	0	0	0	0	0	0	40	0	0	0	0	0	0
Sad (Jump)	0	0	0	0	0	0	0	0	0	39	0	0	1	0	0
Sad (sit)	1	0	0	1	0	0	0	0	0	0	38	0	0	0	0
Sad (Walk)	0	0	0	0	1	0	0	1	0	0	1	37	0	0	0
UTW (Jump)	1	0	0	1	1	0	0	1	0	0	0	0	35	0	1
UTW (sit)	0	0	1	0	0	1	1	0	0	2	0	0	0	35	0
UTW (Walk)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	40

Table II: Confusion Matrix for SVM (*k* means cluster, k=5)

Class	AJ	AS	AW	FJ	FS	FW	HJ	HS	HW	SJ	SS	SW	UJ	US	UW
Angry (Jump)	39	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Angry (Sit)	0	38	0	0	0	0	0	1	0	0	1	0	0	0	0
Angry (Walk)	0	0	37	0	1	0	0	0	1	0	0	1	0	0	0
Fearful (Jump)	0	0	0	40	0	0	0	0	0	0	0	0	0	0	0
Fearful (Sit)	1	0	1	1	35	0	0	1	0	1	0	0	0	0	0
Fearful (Walk)	0	0	0	0	0	39	0	0	0	0	0	0	0	0	1
Happy (Jump)	0	0	1	0	0	0	38	0	0	0	0	1	0	0	0
Happy (Sit)	0	0	0	0	0	0	0	40	0	0	0	0	0	0	0
Happy (Walk)	0	0	0	0	0	0	0	0	40	0	0	0	0	0	0
Sad (Jump)	0	0	0	0	0	0	0	0	0	39	0	0	1	0	0
Sad (sit)	1	0	0	1	0	0	0	0	0	0	38	0	0	0	0

Sad (Walk)	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**37**	**0**	**0**	**0**
UTW (Jump)	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**37**	**0**	**0**
UTW (sit)	**0**	**0**	**1**	**0**	**0**	**1**	**1**	**0**	**0**	**2**	**0**	**0**	**0**	**35**	**0**
UTW (Walk)	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**40**

Table III: Confusion Matrix for KNN (*k* means cluster, k=3)

Class	**AJ**	**AS**	**AW**	**FJ**	**FS**	**FW**	**HJ**	**HS**	**HW**	**SJ**	**SS**	**SW**	**UJ**	**US**	**UW**
Angry (Jump)	**37**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**
Angry (Sit)	**0**	**40**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Angry (Walk)	**0**	**0**	**38**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**
Fearful (Jump)	**0**	**0**	**0**	**40**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Fearful (Sit)	**1**	**0**	**1**	**0**	**37**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Fearful (Walk)	**0**	**0**	**0**	**0**	**0**	**39**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**
Happy (Jump)	**0**	**0**	**0**	**0**	**0**	**0**	**39**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**
Happy (Sit)	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**40**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Happy (Walk)	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**38**	**0**	**1**	**0**	**0**	**0**	**0**
Sad (Jump)	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**40**	**0**	**0**	**1**	**0**	**0**
Sad (sit)	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**38**	**0**	**0**	**0**	**0**
Sad (Walk)	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**37**	**0**	**0**	**0**
UTW (Jump)	**1**	**0**	**0**	**1**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**35**	**0**	**1**
UTW (sit)	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**37**	**0**
UTW (Walk)	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**39**

Table IV: Confusion Matrix for KNN (*k* means cluster, k=5)

Class	**AJ**	**AS**	**AW**	**FJ**	**FS**	**FW**	**HJ**	**HS**	**HW**	**SJ**	**SS**	**SW**	**UJ**	**US**	**UW**
Angry (Jump)	**35**	**0**	**0**	**1**	**1**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**0**
Angry (Sit)	**0**	**39**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Angry (Walk)	**0**	**0**	**36**	**0**	**1**	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**
Fearful (Jump)	**0**	**0**	**0**	**39**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Fearful (Sit)	**0**	**0**	**0**	**0**	**40**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Fearful (Walk)	**0**	**0**	**0**	**0**	**0**	**39**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**
Happy (Jump)	**1**	**0**	**1**	**0**	**0**	**0**	**37**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**
Happy (Sit)	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**39**	**0**	**1**	**0**	**0**	**0**	**0**	**0**
Happy (Walk)	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**40**	**0**	**0**	**0**	**0**	**0**	**0**
Sad (Jump)	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**40**	**0**	**0**	**1**	**0**	**0**
Sad (sit)	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**38**	**0**	**0**	**0**	**0**
Sad (Walk)	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**37**	**0**	**0**	**0**
UTW (Jump)	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**38**	**0**	**1**
UTW (sit)	**0**	**0**	**1**	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**35**	**1**
UTW (Walk)	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**38**

Table V: Performance measure observed with *k*-means cluster k=3 and k=5 of SVM

Emotion/ Action	**Precision (%)**		**Recall (%)**		**Specificity (%)**		**F-Measure (%)**	
	k=3	**k=5**	**k=3**	**k=5**	**k=3**	**k=5**	**k=3**	**k=5**
Angry (Jump)	50	55.2	100	95.1	87.5	100	66.7	65
Angry (Sit)	66.7	62.8	80	91.6	95	90	72.7	65
Angry (Walk)	60	62.1	60	91.6	95	90	60	52.4
Fearful (Jump)	100	89.2	80	91.2	100	90	33.3	58.2
Fearful (Sit)	66.7	64.5	40	45.1	97.5	87.6	50	30
Fearful (Walk)	100	88	40	50	100	79.5	57.1	30
Happy (Jump)	80	91.1	80	61.5	97.5	95.6	80	75.6
Happy (Sit)	66.7	65	80	61.5	95	95.6	72.7	94.2
Happy (Walk)	71.4	70	100	90.7	95	90	83.3	82.2
Sad (Jump)	50	55.2	100	95.1	87.5	100	66.7	65
Sad (sit)	100	88	40	50	100	79.5	57.1	30
Sad (Walk)	80	91.1	80	61.5	97.5	95.6	80	75.6
UTW (Jump)	66.7	62.8	80	91.6	95	90	72.7	65
UTW (sit)	60	62.1	60	91.6	95	90	60	52.4
UTW (Walk)	100	89.2	80	91.2	100	90	33.3	58.2
Average	**79.5**	**80.9**	**76.7**	**79.9**	**95.8**	**97.5**	**91.7**	**92.4**

VI. Conclusion And Outlook

The techniques for recognizing emotion from video using spatio temporal interest points are presented this paper. In this experiment, five emotions (angry, fear, happy, sad and untrustworthy) and three actions (jumping, sitting, and walking) of 10 persons (male and female) are used for experimental purpose. The KNN and multiclass-SVM classifier with polynomial and RBF kernals are used to evaluate the performance of the feature in the video sequences. 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, 2007.
[2]. R. Poppe, "A survey on vision-based human action recognition" IVC 28 , 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", 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, 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" Proceedings of IEEE International Conferences on Computer Vision, pp. 1395–1402, 2005
[17]. Eum, Hyukmin, et al. "Continuous Human Action Recognition Using Depth-MHI-HOG and a Spotter Model" Sensors, 5197-5227, 2015
[18]. J.R. Quinlan," 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, "Overview on Emotion Recognition System" International Conference on Soft-Computing and Network Security, Coimbatore, Feb. 25 – 27, 2015.
Stefano Piana, "Real-time automatic emotion recognition from body gestures" Human-Computer Interaction (cs.HC); Computer Vision and Pattern Recognition, 2014.
[21]. H. Bay, T. Tuytelaars, and L. Van Gool "SURF: Speeded Up Robust Features", IEEE Trans. Pattern Anal. Mach. Intell. 23(3), 257–267, 2007.
[22]. R.Santhoshkumar, M.Kalaiselvi Geetha, J.Arunnehru, "SVM – KNN based Emotion Recognition of Human in Video using HOG Feature and KLT Tracking Algorithm", International Journal of Pure and Applied Mathematics, Volume 117 No. 15, 621-634, 2017.
[23]. R.Santhoshkumar, M.Kalaiselvi Geetha, J.Arunnehru, "Activity Based Human Emotion Recognition in video", International Journal of Pure and Applied Mathematics, Volume 117 No. 15, 1185-1194, 2017.