ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5,Special Issue 12, April 2018

Automated Anger Detection from Human Activity in Video

V.ABINAYA[1], Dr. M. KALAISELVI GEETHA[2], R.SANTHOSH KUMAR [3]

PG Scholar, Department of CSE, Annamalai University, Chidambaram, India [1]
Associate Professor, Department of CSE, Annamalai University, Chidambaram, India [2]
PhD Scholar, Department of CSE, Annamalai University, Chidambaram, India [3]

Abstract: Recognizing emotions from human full body activity is attracting the researchers in recent days due to its potential applications. Detecting the suspicious activity and thereby preventing any terrorism focuses on emotion recognition. This work aims to detect emotions from actions particularly distinguishing anger from other emotions in video. Static and dynamic features are extracted from video frames. The static feature extraction include Histogram of oriented gradients (HOG) feature extraction, whereas the dynamic feature extraction include calculating distance and speed between pixels in frames. That extracted feature is fed into Support vector machine (SVM*)* classifier. The highest F-measure value obtained by SVM classifier is 95.04%. GEMEP dataset is utilized for emotion detection.

Keywords: Emotion Recognition, suspicious activity, anger detection, Histogram of oriented gradients (HOG), Support vector machine (SVM).

I. Introduction

Emotion recognition from body movements can be used in various application areas. The importance of emotion recognization is to avoid or protect the public or a person before something goes wrong. The emotions of the humans can be detected through their activities. Various emotions like happy, anger, sad, fear etc., may be identified from human activity. The aim of this work is to distinguish anger and non-anger emotions from human activities. In current scenario, it is vital to monitor the public areas. Monitoring includes protecting the public from any suspicious activity. Suspicious activity detection can be done by identifying the human emotions through videos. In this work, the distance, speed and HOG feature extraction is used [1]. Then the extracted feature is fed into SVM classifier to get the accuracy percentage of emotion recognized.

A. Outline of the work

This paper deals with human emotion (anger or non-anger) recognition, which aims to identify emotion from the video sequences. The proposed approach is evaluated using GEMEP dataset. The videos are converted into frames. Thus, the features are extracted using HOG feature extraction, distance & speed feature. The extracted features are modeled by SVM classifiers for training and testing. The rest of the paper is structured as follows. Section 2 gives discussion about related works. Section 3 provides an overview of the proposed system. Section 4 describes experimental results. Finally, Section 5 Gives the contributions made by this current work and concludes with directions for future research based on my results.

II. Literature survey

Radoslaw Niewiadomski et al. [1] recognized that laughter from the human full-body movement in social and ecological contexts. A vision-based system prototype for automated laughter detection from human full-body movement was designed and evaluated. Jing Zhu et al. [2] concluded that the concept of emotion deserves a more distinctive and central place in philosophical theories of action. Nesrine Fourati et al. [3] conveyed they present the Random Forest based feature selection approach for the identification of relevant expressive body cues in the context of emotional body expression classification and also discuss the ranking of relevant body cues according to each expressed emotion across a set of daily actions.Neha Shirbhate et al. [4] They are proposing a system that uses semantic rules to define emotional activities. We have opted for semantics-based approach instead of machine learning

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5,Special Issue 12, April 2018

enables us to detect the actions without requiring to train the system. This also makes the system better performance-wise; and enables action detection in real time. Bruce D. Keefe et al. [5] presented a database of high-definition videos for the study of traits inferred from whole-body actions. The database they can be used are useful database of stimuli of multiple actions conveying multiple traits to investigate the contribution of static and dynamic body postures to the perception of body actions. They discuss potential uses for the database in the analysis of the perception of whole-body actions. Haris Zacharatos et al. [6] outlines the findings relating to the area of body emotion recognition by describing emerging

techniques and modalities as well as recent advances on the challenging task of automatic emotion recognition. Important aspects are analyzed, application areas and notation systems are described and the importance for movement segmentation is discussed. The survey concludes with a detailed discussion on unsolved problems in the area and provides promising directions for future work. Nele Dael et al. [7] they investigated to what extent these expression patterns support explicit or implicit predictions from basic emotion theory, bi-dimensional theory, and componential appraisal theory. Konrad Schindlera et al. [8] approaches recognition of basic emotional categories from a computational perspective. They construct a biologically plausible hierarchy of neural detectors, which can discriminate seven basic emotional states from static views of associated body poses. Michael Garber-Barron et al. [9] explored the use of features that represent body posture and movement for automatically detecting people's emotions in non-acted standing scenarios. We focused on four emotions that are often observed when people are playing videogames: triumph, frustration, defeat and concentration. Castellano et al. [10] presented an approach for the recognition of acted emotional states based on the analysis of body movement and gesture expressivity. According to research showing that distinct emotions are often associated with different qualities of body movement, we use non-propositional movement qualities to infer emotions, rather than trying to recognise different gesture shapes expressing specific emotions. J. Arunnehru et al. [11] proposed an Automatic human emotion Recogniton in surveillance video based on gesture dynamic's features and the features are evaluated by svm , Naïve Bayes and dynamic time wrapping. M. KalaiselviGeetha et al. [12] developed a video retrievel applications for video classification and shot detection using Block intensity comparison code (BICC) and unsuperized shot detection. A Noval AANN misclusting Rate(AMR) algorithm is used to detect the shot transitions. J. Arunnehru et al. [13] assigns motion intensity code for action recognition in surveillance video using Region Of Interest (ROI) from the difference image. J. ArunNehru et al. [14] proposed an applications for Action Recognition in automated surveillance. The 18 dimensional Block intensity vector are extracted and evaluated through SVM.

III. PROPOSED SYSTEM

A. Block Diagram

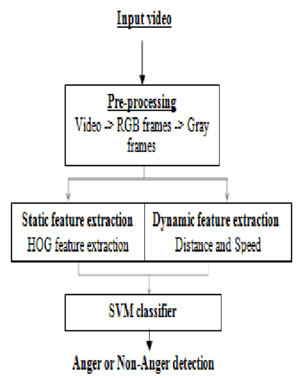


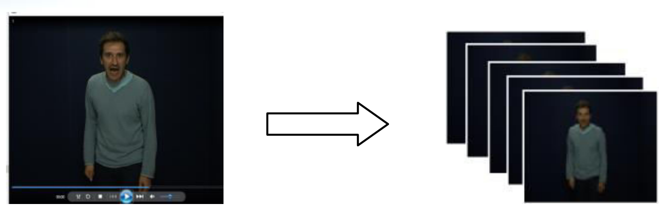
Fig. 1 Block diagram for proposed work.

Fig 1. shows that the steps involved in this proposed approach, First step in the proposed system is preprocessing, Secondly the feature is extracted and finally, the extracted feature is send through Classifier.

B. Preprocessing

Pre-processing is a common name for operations with images at the lowest level of abstraction -- both input and output are intensity images. The goal of pre-processing is making to suppresses unwanted distortions or enhances some image features important for next processing stage.

Here the steps involved in preprocessing are:



2(a) VIDEO

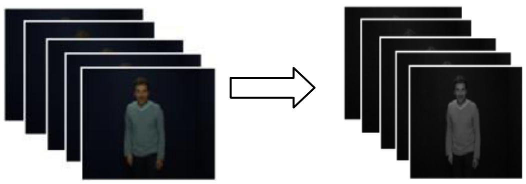
2(b) RGB Frames

Fig.2. Converting videos into RGB frames.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5,Special Issue 12, April 2018

Here 2(a) shows the video and 2(b) shows the RGB frames, Thus the fig 2, shows that the step1 in preprocessing.



3(a) RGB Frames — 3(b) GRAY Frames

Fig.3. Converting RGB frames into gray frames.

Here 3(a) shows the RGB frames and 3(b) shows the GRAY frames, Thus the fig 3 shows that the step2 in preprocessing.

4(a) GRAY Frames — 4(b) Bounding box

Fig.4. Converting Gray frame into Bounding box frame

Here 4(a) shows the GRAY frame and 4(b) shows the Bounding box frame, Thus the fig 4 shows that the step3 in preprocessing.

C. Static feature extraction : Histogram of oriented gradients (HOG)

Histograms are collected totals of data prepared into a set of predefined bins. Imagine that a matrix contains information of an image intensity value in the range 1 to 256. The histogram of an image is a design of the gray-level values vs. the number of pixels at that value. The feature will be used then for classification of emotion and action recognition. The HoG feature process first calculates the gradient of the input image. A typical method is to apply a one-dimensional discrete differential mask as the horizontal orientation (D_k = [–1 0 1]) and vertical orientation.

$$D_y = [-101]^T \quad \text{.......(1)}$$

Convolution mask of horizontal orientation:

$$I_X = H(x, y, t) * D_X \quad \text{.......(2)}$$

Convolution mask of vertical orientation:

$$I_Y = H(x, y, t) * D_Y \quad \text{.......(3)}$$

Size of gradient:

$$|G| = \sqrt{I_X^2 + I_Y^2} \quad \text{.......(4)}$$

Orientation of gradient:

$$\theta = \arctan \frac{I_Y}{I_X} \quad \text{.......(5)}$$

Signed gradient:

$$\propto_{signed} = \begin{cases} \propto & \propto \geq 0 \\ \propto + 360 & \propto < 0 \end{cases} \quad \text{.......(6)}$$

Unsigned gradient:

$$\propto_{signed} = \begin{cases} \propto & \propto \geq 0 \\ \propto + 180 & \propto < 0 \end{cases} \quad \text{.......(7)}$$

When the image is created, convolution masks of the horizontal and vertical orientations (Equations (2) and (3), respectively) are applied to the image, and the orientation and slope size are calculated. Second, calculate histograms of the cells are divided. The value of each pixel in the cell is calculated as the gradient of the orientation through an advanced gradient calculation.

These orientation histogram bands are spread on values, which are set as the number of bins. Rectangular shapes in the image of comprised are cells. As an expression of the gradient, the histogram bands are evenly distributed from 0 to 360 degrees (Equation (6)) or from 0 to 180 degrees (Equation (7)). HOG features to extract the bounded box frames are obtained. Then image is divided into 3 * 3 cells, where each cell consists of 6 bin histogram , in order to obtain the HOG features.

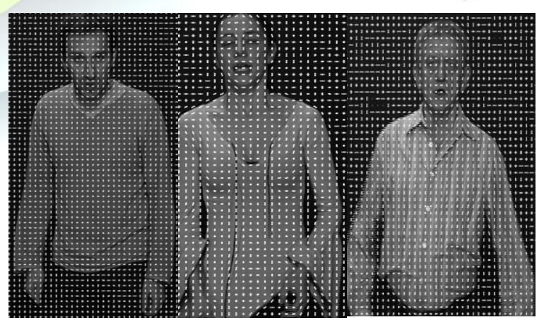


Fig.5.HOG Feature Extraction

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5,Special Issue 12, April 2018

D. Dynamic feature extraction

Centroid

The centroid is also called as geometric center of a plane figure is the arithmetic mean position of all the points in the shape. The definition extends to any object in n-dimensional space: its centroid is the mean position of all the points in all of the coordinate directions.

Where (x1,y1) and (x2,y2) are two coordinates points. The two points are taken from pixels in two frames. The two points are used to calculate Distance and Speed.

Distance

To calculate the Euclidean distance D, the difference image found by subtracting by two consecutive frames [1].
The Euclidean Distance Between two frames

$$D = \text{sqrt}((x2 - x1)^2 + (y2 - y1)^2); \quad \ldots\ldots.(8)$$

Speed

To calculate the speed S, dividing the distance of the frame D by time T [1]. Where 0.08 is the time difference between two frames.
The average speed of the motion between n frames

$$S = \frac{D}{0.08}; \quad \ldots\ldots.(9)$$

E. SVM classifier

Support vector machine (SVM) is based on the principle of structural risk minimization (SRM). Like RBFNN, support vector machines can be used for pattern classification and nonlinear regression [8]. SVM constructs a linear model to estimate the decision function using non-linear class boundaries based on support vectors.

If the data are linearly separated, SVM trains linear machines for an optimal hyperplane that separates the data without error and into the maximum distance between the hyperplane and the closest training points. The training points that are closest to the optimal separating hyperplane are called support vectors.

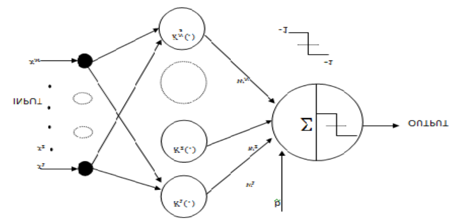


Fig.6. Architecture of the SVM

Fig 6. shows the architecture of SVM, SVM maps the input patterns into a higher dimensional feature space through some nonlinear mapping chosen a priori. A linear decision surface is then constructed in this high dimensional feature space. Thus, SVM is a linear classifier in the parameter space, but it becomes a nonlinear classifier as a result of the nonlinear mapping of the space of the input patterns into the high dimensional feature space. In this work two types of feature extraction is used static feature extraction and dynamic feature extraction. The static feature extraction is HOG feature extraction and the dynamic feature extraction is DISTANCE AND SPEED calculation.

In SVM, both types of feature extractions are fed and the output is predicted. The higher accuracy level obtained in SVM classifier is 95.04% from HOG Feature extraction in BIN 14. Fig 7, shows the confusion matrix and the predicted emotion using SVM classifier.

Similarly, Fig 8. shows the confusion matrix and the predicted emotion for Dynamic Feature extraction (DISTANCE,SPEED) using SVM classifier.

SVM classifier(HOG feature extraction)

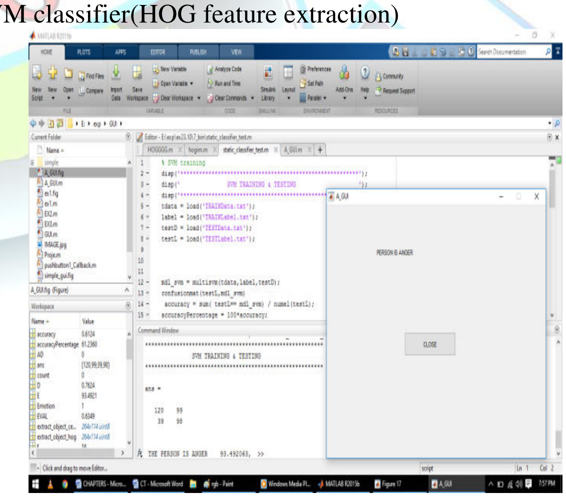


Fig 7. SVM Classifier(HOG Feature extraction)

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5,Special Issue 12, April 2018

SVM classifier(distance & speed)

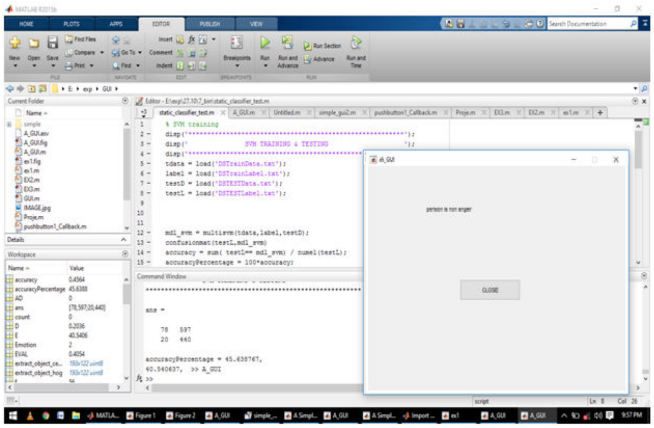


Fig 8. SVM Classifier(DISTANCE & SPEED)

IV. Experinmental Results

In this work, The GEMEP dataset is used. Hence, HOG is used for feature extraction . That extracted feature is send through SVM classifier to get the accuracy level. Thus different dimensional values are calculated by different BIN's used to predict most accurate value is taken. The highest F-measure value obtained by SVM classifier is 95.04%.

A. Dataset

A data set (or dataset, although this spelling is not present in many contemporary dictionaries like Merriam-Webster) is a collection of data. Most commonly a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. The data set may comprise data for one or more members, corresponding to the number of rows.

The term data set may also be used more loosely, to refer to the data in a collection of closely related tables, corresponding to a particular experiment or event.

B. GEMEP Dataset

The GEneva Multimodal Emotion Portrayals (GEMEP) is a collection of audio and video recordings featuring 10 actors portraying 18 affective states, with different verbal contents and different modes of expression. It was created in Geneva by Klaus Scherer and Tanja Bänziger, in the framework of a project funded by the Swiss National Science Foundation (FNRS 101411-100367) and with support of the European Network of Excellence "Humaine" (IST-2002-2.3.1.6 Multimodal Interfaces, Contract no. 507422). Fig 9. shows the example pictures of actors performing various emotions.

Fig 9. GEMEP Dataset

C. Confusion matrix

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. The confusion matrix itself is relatively simple to understand, but the related terminology can be confusing. The most basic terms, which are whole numbers (not rates):

- True Positives (TP) - These are the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also yes.

- True Negatives (TN) - These are the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5,Special Issue 12, April 2018

- False Positives (FP) – When actual class is no and predicted class is yes.
- False Negatives (FN) – When actual class is yes but predicted class in no.

TABLE 1.
Confusion Matrix

ACTUAL	PREDICTED	
	POSITIVE	NEGATIVE
POSITIVE	TP	TN
NEGATIVE	FP	FN

The above table represents the confusion matrix, TP,TN,FP,FN are actual and predicted values.

- Precision

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$\text{Precision} = \frac{TP}{TP+FP} \quad ....(10)$$

- Recall

Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes.

$$\text{Recall} = \frac{TP}{TP+FN} \quad ....(11)$$

- F-Measure

In statistical analysis of binary classification, the F-score (also F-score or F-measure) is a measure of a test's accuracy. It considers both the precision and the recall of the test to compute the score.

$$\text{F-score}=2.\frac{PRECISION.RECALL}{PRECISION+RECALL} \quad ....(12)$$

D. Performance measures

TABLE.2.
Various accuracy level using SVM classifier

S.NO	BIN (B)	DIMENSION VALUE (x*y*B)=(3*3*B)	ACCURACY %
1.	7	63	89.04
2.	8	72	88.76
3.	9	81	90.52
4.	10	90	86.80
5.	12	108	89.04
6.	14	126	95.04

In Table 2. describes various accuracy level predicted using different dimensional value and also different bin's used to calculate dimensional values are listed.

The graphical representation of various accurate levels predicted and shown below.

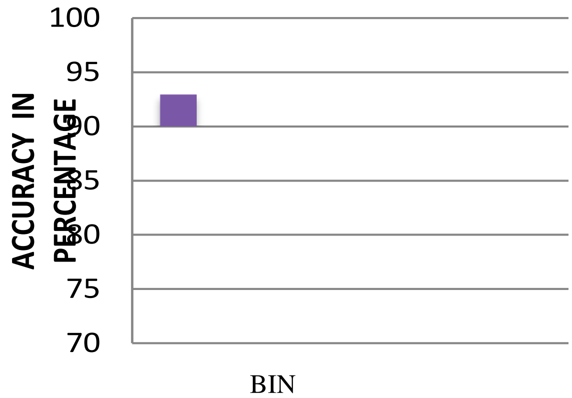


Fig.10. Chart for different accuracy level from SVM classifier

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5,Special Issue 12, April 2018

E. Results obtained with confusion matrix

TABLE.3.

Confusion Matrix obtained for various bin's

		7-BIN		8-BIN		9-BIN		10-BIN		12-BIN		14-BIN	
		ANGER	NON-ANGER	ANGER	NON-ANGER	ANGER	NON-ANGER	ANGER	NON-ANGER	ANGER	NON-ANGER	ANGER	NON-ANGER
7-BIN	ANGER	219	0	-	-	-	-	-	-	-	-	-	-
	NON-ANGER	39	98	-	-	-	-	-	-	-	-	-	-
8-BIN	ANGER	-	-	218	1	-	-	-	-	-	-	-	-
	NON-ANGER	-	-	39	98	-	-	-	-	-	-	-	-
9-BIN	ANGER	-	-	-	-	215	4	-	-	-	-	-	-
	NON-ANGER	-	-	-	-	8	129	-	-	-	-	-	-
10-BIN	ANGER	-	-	-	-	-	-	360	0	-	-	-	-
	NON-ANGER	-	-	-	-	-	-	78	151	-	-	-	-
12-BIN	ANGER	-	-	-	-	-	-	-	-	217	2	-	-
	NON-ANGER	-	-	-	-	-	-	-	-	39	98	-	-
14-BIN	ANGER	-	-	-	-	-	-	-	-	-	-	215	4
	NON-ANGER	-	-	-	-	-	-	-	-	-	-	35	102

TABLE 4.

Precision, Recall, F-Score obtained for various bin's

BIN'S	PRECISION (IN %)	RECALL (IN %)	F-SCORE (IN %)
7	84.88	100.00	91.82
8	84.82	99.54	91.59
9	85.02	100.00	92.05
10	82.27	100.00	90.27
12	84.88	100.00	91.82
14	91.16	100.00	95.60

V. CONCLUSION AND FUTURE WORK

A. Conclusion

In this proposed work, The techniques to detect anger or non anger emotions through full body movement activity. The emotion can be detected through static feature extraction (HOG feature extraction) & dynamic feature extraction (distance and speed) and SVM classifier to predict correct accuracy level. The performance can be evaluated by recall, precision and F-Measure. The system is used in various application to detect anger of the person to avoid critical situation.

The idea is to make the user explicitly aware that a machine is analyzing his or her body cues and hence create a kind of dialogue between the user's emotional behaviour and the machine's appropriate responses. This kind of situation is much more similar to person-person communication. The kinds of emotional signals that get sent during these kinds of interactions may be more intentional, very different and potentially easier to analyze by statistical algorithms.

B. Future work

Our proposed work is detecting and displaying whether the person is anger or non-anger emotion, Our feature work aims to detect all emotions of the human through their activity and displaying the emotion. The future work is helpful in person-person interactions.

ACKNOWLEDGMENT

I express my deep sense of gratitude to my guide Dr. M. KALAISELVI GEETHA, M.Tech., Ph.D., Associate professor, Department of Computer Science and Engineering for her persistent interest in this project and her guidance and constant encouragement during this project without which the project would not have been materialized.

I would like to express my sincere gratitude and love to my parents and friends for their support and encouragement for the successful completion of the project.

REFERENCES

[1] Radoslaw Niewiadomski, Maurizio Mancini, Giovanna Varni, Gualtiero Volpe, and Antonio Camurri,." Automated Laughter Detection From Full-Body Movements",. IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS, VOL. 46, NO. 1, FEBRUARY 2016.

[2] Jing Zhu & Paul Thagard. "Emotion And Action"., Philosophical Psychology, Vol. *15, No. 1, 2002 ISSN 0951-5089.

[3] Nesrine Fourati, Catherine Pelachaud . "Relevant body cues for the classification of emotional body expression in daily actions",8/15/$31.00 ©2015 IEEE.

[4] Neha Shirbhate ,Kiran Talele, Mumbai, India. "Human Body Language Understanding for Action Detection using Geometric Features"1/16/$31.00,2016, IEEE.

[5] Bruce D. Keefe & Matthias Villing & Chris Racey & Samantha L. Strong & Joanna Wincenciak & Nick E. Barraclough.," A database of whole-body action videos for the study of action, emotion, and untrustworthiness".

[6] Haris Zacharatos, Christos Gatzoulis and Yiorgos Chrysanthou "Automatic Emotion Recognition based on Body Movement Analysis:A Survey".

[7] Nele Dael, Marcello Mortillaro, and Klaus R. Scherer University of Geneva.," Emotion Expression in Body Action and Posture"., Emotion © 2011 American Psychological Association 2012, Vol. 12, No. 5, 1085–1101.

[8] Konrad Schindlera, Luc Van Goola,b, Beatrice de Gelderc.," Recognizing Emotions Expressed by Body Pose: a Biologically Inspired Neural Model"

[9] Michael Garber-Barron , Mei Si .,"Using Body Movement and Posture for Emotion Detection in Non-Acted Scenarios"., WCCI

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5,Special Issue 12, April 2018

2012 IEEE World Congress on Computational Intelligence June, 10-15, 2012 - Brisbane, Australia.

[10] Ginevra Castellano, Santiago D. Villalba, and Antonio Camurri., Infomus Lab, DIST, University of Geno MLG, School of Computer Science and Informatics, University College Dublin., "Recognising Human Emotions from Body Movement and Gesture Dynamics".

[11] J. Arunnehru and Kalaiselvi M 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.

[12] 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.

[13] 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.

[14] Arunnehru, J., 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.

BIOGRAPHY

First Author V.ABINAYA received her B.E Computer Science and Engineering from Annamalai University, India, in the year 2016; She is doing her M.E Computer Science and Engineering at Annamalai University,India.

Second Author Dr. M. KALAISELVI GEETHA gained her Ph. D in Computer Science and Engineering from Annamalai University, India; M.Tech in Computer Technology from I.I.T. Delhi, India under Quality Improvement Programme (QIP); B.E (CSE) from Institute of Road and Transport Technology, Erode, India. She joined Annamalai University as Lecturer in 1994 and currently serving as Associate Professor in Dept. of Computer Science and Engineering in Annamalai University. She is life member in CSI and ISTE professional bodies. She has authored over 70 national and international journal papers to her credit. Also, she has published many chapters in different books published by International publishers. In addition to this, she has also edited a book published by International publisher. She has successfully completed UGC sanctioned research project worth Rs. 13 lakhs on video surveillance. Five research scholars have completed their doctoral research under her supervision. Currently 2 scholars are pursuing their doctoral research under her supervision. Her research interests include Video Processing and Video Surveillance, Pattern Classification and Artificial Intelligence.

Third Author R. Santhoshkumar Received his B.E. degree in Information Technology and M.E. degree in Computer Science and Engineering from Annamalai University, Annamalainagar, Tamilnadu. India. He is currently pursuing his Ph.D. in Computer Science and Engineering at Annamalai University India. He has published 3 research papers in international journals and conferences. His current research interests include computer vision, video processing, pattern recognition, machine learning and deep learning.