IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-6, June 2019 E-ISSN: 2347-2693

Multitask sparse Learning based Facial Expression Classification

Pratik Nimbal[1*], Gopal Krishna Shyam[2]

[1,2] Department of Computer Science, Reva University, Bangalore, India

Corresponding Author: pratiknimbal5@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i6.197202 | Available online at: www.ijcseonline.org

Accepted: 11/Jun/2019, Published: 30/Jun/2019

***Abstract*—** In today's era, Facial Expression and recognition is a very challenging and fascinating subject with regards to field of AI and pattern recognition because of developmental psychology and human machine interface. For outward appearance designing classifiers with high reliability is a significant advance in this research. This paper represents a framework for person dependent expressions by combining all types of facial recognition types of facial by means of various multiple kernel learning in Support vector machines(SVM). We contemplated the impact of MKL for learning the piece loads and observationally assess the aftereffects of six fundamental expressions with impartial expression. included. In our investigations we have joined two mainstream facial element portrayals, dlib library and Multikernel SVM with polynomial kernel. Our experimental results on the cohn-Kanade face database as well as manually included database demonstrate that this framework out performs the state-of-arts, conventional techniques and straightforward MKL based multiclass SVM for facial expression recognition.

***Keywords*—**Facial Expression Recognition, Multikernel, Support Vector Machines.

I. INTRODUCTION

The human face gives various social sign which are basics for relational correspondence in our regular day to day existence. The human face additionally holds significant amount of traits and data about the individual, for example, outward appearance, ethnic, gender, and age. Outward appearance is a development of facial muscles by a human automatically when they feel something like anger, bliss and fear… . and so on (figure 1). People can perceive outward appearances for all intents and purposes immediately. Be that as it may, solid and completely computerized articulation by fully automated speech is still a different and difficult task . Different methodologies have just been endeavoured towards tending to this issue, however the complexities included by conditions like between close to home variety (for example gender, ethnic) and irregularity of procurement conditions (I e. illumination, resolution) have made the assignment very confused and testing. It is trusted that the mechanized examination of outward appearances can encourage machine view of human facial conduct, and in this way opens up the method for bringing outward appearances into human-machine collaboration as another methodology towards making the process progressively regular and productive. It can likewise empower the grouping and evaluation of physical appearances generally available for the exploration in conduct science and medication via computerized mental perception of people. Keeping these in thought, this postulation tends to the different complexities identified with the characterization of the experienced outward appearances present in static facial pictures and accordingly gives an answer of the issue of grouping the seven significant outward appearances to be specific, unbiased, outrage sicken, fear, joy, trouble and amazement. Six of these articulations with the exception of the nonpartisan have been characterized as the essential human feelings by Ekman.

It can improve the correspondences among human and machines, and it is helpful in human- machine cooperation, in this manner later on, robots/machines can comprehend human conduct. Further applications lie in security, driver wellbeing and sociologies as a device to investigate human full of feeling conduct. The illustration of six different emotions is described below:

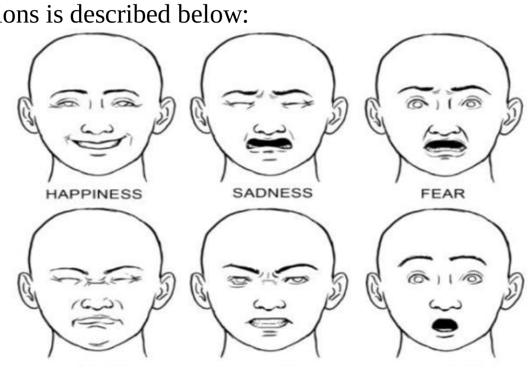

Figure 1 : Six different expressions

Here in this research, we make use of dlib library for frontal face detector mainly and widely used for face detection. This algorithm training is slow but detection is very fast here. Six basic emotions are trained here for facial expression recognition, they are anger , fear , happy or smile, ,surprise , disguise and sadness. Neutral is considered when all are not detected and negative result is obtained Section I contains the introduction of Facial Expression Recognition , Section II contain the related work of Facial Expression recognition based on action units and multitask sparse learning . Section III contain the architecture and pipeline implementation part of this research paper. Section IV contain the experimental results compared with methods explained. Section V includes conclusion and future scope.

II. RELATED WORK

A. Facial Expression Analysis

1) Most outward appearance examination strategies for the most part pursue the previously mentioned two classes: AU-based and message and sign judgment techniques. Despite the fact that this paper has a place with the last class, it is as yet important to give a finished audit on related takes a shot at expression recognition. AU-based outward appearance examination propelled by the outstanding investigation on facial action, Facial Action Coding System(FACS)[17]. In this system, the subtle changes in facial appearance are encoded into 32 AUs with individual phonetic depiction. Since every fundamental articulations can be disintegrated into a few related AUs, the articulation acknowledgment issue can be moved to AUs discovery issue. Bartlett et al[18]. perceived six single upper face AUs, however no synchronous AUs are considered in mix. Tian et al.[19] distinguished 16 AUs from face image sequences using lip tracking, template matching and neural networks. More works have been done on spontaneous facial articulation information via automatic recognition of AU's[20], [21],[22], [23],[24], [25]. A few contrasts among unconstrained and purposeful facial conduct are additionally contemplated . As of late, AU location and AU-based articulation acknowledgment techniques make a ton of efficient progress.. Sandbach et al. abuse 3-D movement based highlights between casings of 3-D facial geometry sequences for dynamic AU detection and further expression recognition. AU-based methods break down the outward appearances into changed individual muscle exercises, and afterward deduce the articulation classes dependent on the AU location results. These strategies can have extraordinary portrayal control, yet AU recognition itself is very troublesome and it is as yet an open issue to the network.

2) Message and sign judgement facial expression analysis methods as mentioned in previous research papers consist of the two main steps: facial portrayal and articulation acknowledgment.

a) Facial portrayal infers a lot of highlights from unique facial pictures which are visible from basic to all appearances. Various highlights have been connected to either the entire face or explicit face areas to separate the facial appearance changes, for example, Gabor [20],[26],[27] Haar-like highlights, nearby twofold examples (LBPs)[28],[29] . Zafeiriou and Pitas investigated the diagram structures with landmarks to represent the difference among different expressions. In [44], Facial images are equally bifurcated into small regions, and then LBP features are extracted from these exactly weighted sub-areas to the facial appearance. The LBP highlights are demonstrated to be powerful in face recognition , so this paper will likewise use the LBP features with the same sub-region division strategy. Different from their work, they focused on learning the effective sub-regions statically.

b) Expression recognition plans to accurately sort diverse facial portrayals. Bolster vector machine (SVM)[20],[30],[31] is the most mainstream and viable learning strategy in outward appearance acknowledgment. Reference (paper) is the most comparable work to our own, so it will be considered as the gauge. For reasonable examination, this paper will likewise utilize SVM as the order calculation.

Other than these works, there are additionally a few works using the geometric highlights, for example, the area of facial element focuses (corners of the eyes, mouth, and so on.). A few techniques perform outward appearance examination dependent on 3-D face models [33],[34],[35]. More chips away at combination of sound and visual data can be found in [32].

B. MTSL

Sparsity techniques have pulled in much consideration in PC vision, mixed media and restorative picture networks, and have been utilized in numerous applications, for example, face acknowledgment , foundation subtraction , picture explanation and recovery , and shape earlier based division . Numerous calculation are proposed to take care of these issues of sparsity priors, for example, covetous strategies [basis interest (BP) , coordinating interest, symmetrical coordinating interest (OMP) , or L1 standard unwinding and raised streamlining .

MTSL is an inductive exchange AI approach. It intends to gain proficiency with an issue together with some related issues for better execution. MTSL is then structured in for highlight determination, through empowering various indicators from various assignments to have comparable parameter sparsity designs. MTSL likewise acquired a remunerating execution on written by hand character acknowledgment in. Yuan et al. built up a visual characterization calculation by learning the common parts among various portrayal assignments. As of late, Chen et al. given a quicker answer for MTSL issues.

Assume there are T related assignments, and (x^t , y^t), $i = 1,2,...,N_t$ is the training set of task t, where each sample is represented by K -dimensional features, $x_i^t \in R^K$, and $y^t_i \in \{-1,1\}$ indexes x_i^t is negative or positive. w^t is a K-dimensional vector of representation coefficients for task t. All the w^t s are the rows of the matrix $W = [w^t_k]_{t,k}$, while every column of the matrix W is a T-dimensional vector that means the representation coefficients from the kth feature across different tasks, $w_k = [w^1_k, w^2_k, ..., w^T_k]'$. MTSL aims to learn the shared sparse information among all the tasks. The formulation with $L1/L2$ mixed-norm regularization is described after.

In below described formula where $J^t(w^t, x^t, y^t)$ is the cost function of the tth task, λ is a constant to balance the sparsity, and automatic format for $L1$ norm. The regularization term encourages most columns of matrix W to be zero, and the remaining nonzero columns indicate the corresponding features are shared features across all the tasks.

$$\arg\min_{W} \sum_{t=1}^{T} \frac{1}{N_t} \sum_{i=1}^{N_t} J^t\left(w^t, x_i^t, y_i^t\right) + \lambda \sum_{k=1}^{K} \|w_k\|_2$$

III. IMPLEMENTATION

The proposed system consists of many stages which includes collecting database manually and from cohn-Kanade database, In this section , we make use of multiscale facial appearance representation method and then learning procedures of common and specific patches which are illustrated at each level. At last, we design the classifier with these learned effective patches.

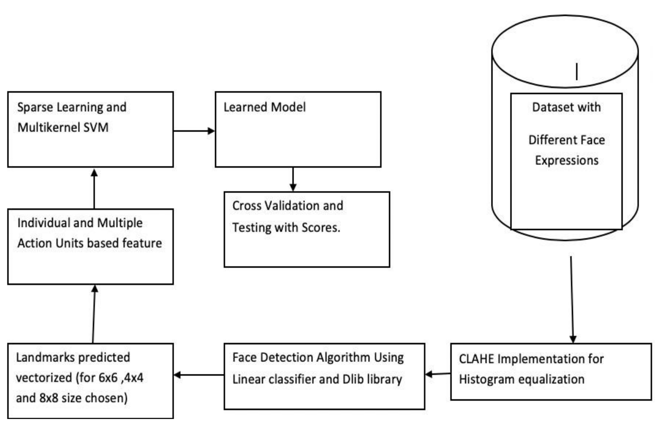

Figure 2: Architecture of implemented system

1) Database Operation

In this step, almost around hundreds of images are collected from cohn-Kanade data base which include images which are equally divided according to six basic expressions mentioned. These images are shuffled and are divided for about 80% and 20% is allocated for training respectively.

2) Pre-processing Stage.

Initially the images are collected, so the first step is to detect the correct position of the image and action units are imposed on the face. The basic unit of facial recognition is action unit. Each action unit is related to contraction of one or more facial muscles. When one or more action units are active then a combination of action units occur . There are total of 68 action units which are spread across and stick to the physical features of the face which clearly describe a facial expression

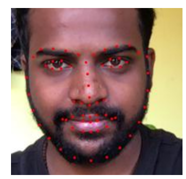

Fig 3: Action units are spread across the physical feature of the face (red dots)

Intra-class is used to compare one expression of same class/expression . For e.g Smile is compared with smile only for distinct results and accuracy.

In Intraclass, each expression is compared with other class to differentiate the values obtained and will be more easier to handpick a particular expression precisely.

Dlib frontal face detector is used to extract the facial feature from the primary camera and the database we have fed the system. The difference after comparing the two expressions is called as convolution value When action units have done their part, landmarks are recognized and vectorized here. After detection of action units, the CLAHE – Contrast limited adaptive histogram equalization is implemented here. While implementing CLAHE, the distinction intensification within the area of a given component or quantity is created by the slope of the modifying or remodeling perform, this can be always proportional to the slope of the adjacent cumulative distribution function (CDF) and so to the worth of the bar chart created for that pixel. CLAHE works because it restricts the intensification by cutting the bar chart or intensity at a particular worth before scheming the CDF. This limits the slope of the CDF and so of the transformation perform. the worth at that the bar chart is clipped, that's known as clip limit, depends on the standardization of the bar chart and so on the scale of the neighbourhood region. Common values restricts the results of amplification to between value of three and four.

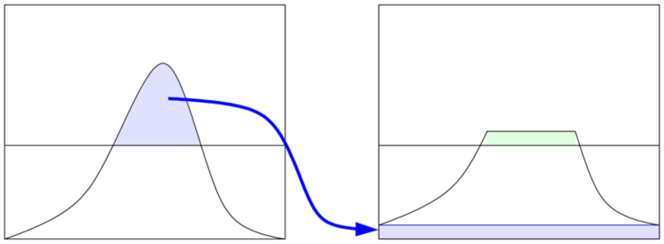

Fig 4 : CLAHE clipping of the higher illumination(slope) of a pixel and limiting it to a equalized value.

In CLAHE, a CDF(Cumulative distributive function) is used to equalize the slope of the histogram and thus reducing the noise and distortion produced in the image and hence this image is used for the next process.

3) Feature Extraction

Feature extraction involves the operation of action units in which they are divided into two classes. Inter-class and Intraclass feature extraction. Both are implemented on a Multikernel SVM.

4) Classification

In this process, training models and prediction models are taken out. The one single expression task is defined here in this process. Each task is a binary expression classification task and are divided into positive samples and negative samples. Here, the data is split into training data and testing data for all classes. Using this both training and testing data, training sets and testing sets are separately done. With training data and training labels, the model is created and the model acts itself as a classifier and is fit in SVM. With validation data and labels, Validation is done and accuracy is predicted here. Here we can find out separate accuracy for single kernel linear SVM and Multikernel Linear SVM

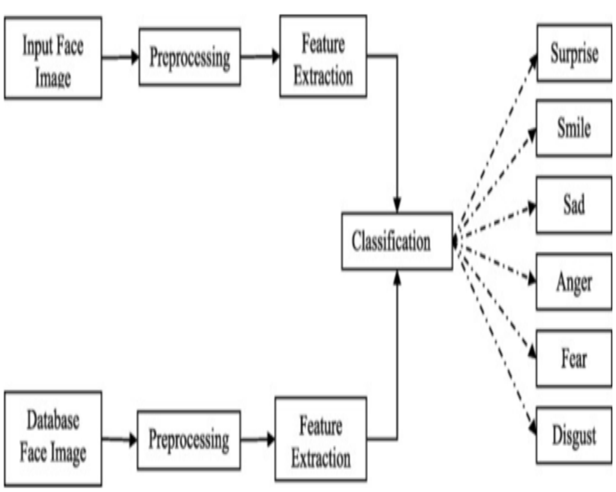

Fig 5: Pipeline of implemented system.

IV. EXPERIMENTAL RESULTS

Fig 6 : Different six kind of expressions detected which are converted to clahe here and displayed on the output console.

Feature vectors generated are stored in form of binary data and arrays and can be viewed ad np.array under Jupyter notebook console, which appears in form of binary data.
The output is displayed on the console of Jupyter notebook as smile , anger, sadness or any specific corresponding expression which is live fed through primary camera of the device.

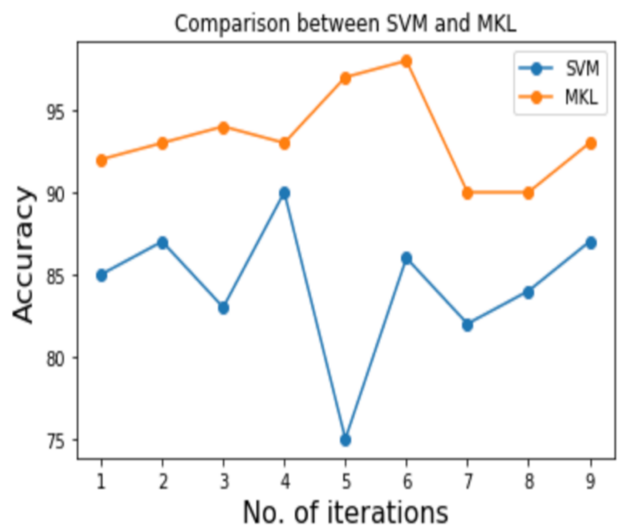

Fig 7 : Comparison between different accuracies obtained.
For calculating accuracies, we have to test the code for 10 times, note down all the values manually and then we can plot a graph for comparing them.

Here, SVM with single kernel graph drops down because
It detects way more time to extract images and feature vector because of single kernel images which are trained in single kernel svm compared to Multikernel svm.

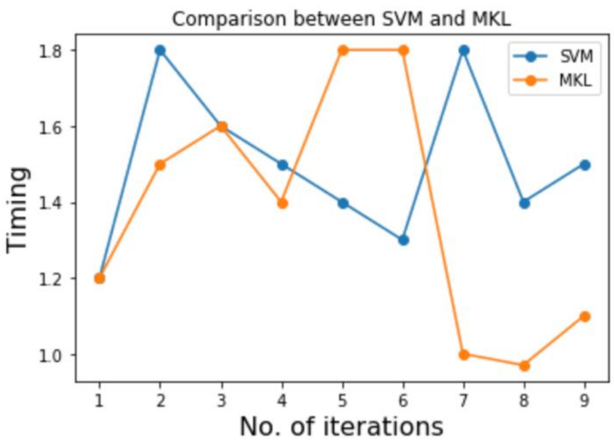

Fig 8: Comparison between Total time taken in SVM training and MKL training.

For calculating total time taken, we have to test the code for 10 times, note down all the values manually and then we can plot the timeseries graph as shown above and then compare them to obtain correct values.

Time taken by Multikernel svm is less compared to linear kernel svm because time taken to train data in Multikernel svm is way far less than linear kernel.

V. Conclusion and Future Scope

Face emotion is difficult to predict precisely because the feature extraction has certain. No stationary characteristics or properties which make it difficult to predict. In this project, we have learned and tuned the weights using several combination of kernels. Algorithm is proposed where solution of mix kernel we have combined kernels and this is showing better behaviour and more predicted accuracy than other base line models in our experiments.

This research throws light on future enhancements that can be carried out. Some of the further enhancements would be implemented to approach and design for parallel computing platform for training Multikernel learning which would help decreases the time required for the research approach.

REFERENCES

[1] Asullah Khalid Alham, Maozhen Li1, Suhel Hammoud and Hao Qi, “Evaluating Machine Learning Techniques for Automatic Image annotation,vol. 11, no. 1, january/february (2009),p.21-235.
[2] O. Marques, N. Barman, "Semi-Automatic Semantic Annotation of Images Using Machine Learning Techniques" Proc. of ISWC(2003), p. 550-565.
[3] J. Liu, B. Wang, M. Li, Z. Li, W. Y. Ma, H. Lu and S. Ma, “Dual Cross-Media Relevance Model for Image Annotation,” in Proceedings of the 15th International Conference on Multimedia(2007), p. 605 – 614.
[4] C. F. Tsai and C. Hung, “Automatically Annotating Images with Keywords: A Review of Image Annotation Systems,” Recent Patents on Computer Science (2008), vol 1, pp 55-68.
[5] Learning Multiscale Active Facial Patches for Expression Analysis Lin Zhong, Qingshan Liu, Peng Yang, Junzhou Huang, and Dimitris N. Metaxas, Senior Member, IEEE
[6] R. Datta, D. Joshi, J. Li and J. Z. Wang, “Image Retrieval: Ideas, Influences, and Trends of the New Age” ACM Computing Surveys (CSUR)(2008), vol. 40,), p. 605 – 614.
[7] L. Cao, J. Luo, H. Kautz and T. S. Huang. “Image Annotation within the Context of Personal Photo Collections Using Hierarchical Event and Scene Models”, In (2009) IEEE Multimedia 11(2), p. 208- 219.
[8] W. Viana, J. B. Filho, J. Gensel, M. Villanova-Oliver and H. Martin, "PhotoMap: From location and time to context-aware photo Annotations", In (2008) Journal of Location Based Services 2(3), p. 211-235.
[9] M. Ames and M. Naaman, “Why We Tag: Motivations for Annotation”. In proc. CHI, ACM Press (2007), p. 971-980
[10] U. WESTERMANN and R. JAIN, "Toward a Common Event Model for Multimedia Applications", In (2007) IEEE Multimedia 14(1), p. 19-29.
[11] M. Davis, N. V. House, J. Towle, S. King, S. Ahern, C. Burgener, Perkel, M. Finn, V.Viswanathan and M. Rothenberg, “MMM2: Mobile Media Metadata for Media Sharing”, Ext. Abstracts CHI (2005), ACM Press, p. 1335-1338.
[12]Tianxia Gong, Shimiao Li and Chew Lim Tan, ”A Semantic Similarity Language Model to Improve Automatic image annotation”, In (2010) 22nd International Conference on Tools with Artificial Intelligence.
[13] Lei Ye, Philip Ogunbona and Jianqiang Wang, "Image Content Annotation Based on Visual Features” Proceedings of the Eighth IEEE International Symposium on Multimedia (ISM'06).
[14] Yunhee Shin, Youngrae Kim and Eun Yi Kim, "Automatic textile image annotation by predicting emotional conceptsfrom visual features”. In (2010) Image and Vision Computing, p. 28.
[15] Ran Li, YaFei Zhang, Zining Lu, Jianjiang Lu and Yulong Tian, “Technique of Image Retrieval based on Multi-label Image Annotation”, In (2010) Second International Conference on MultiMedia and Information Technology.
[16] T. Jiayu, “Automatic Image Annotation and Object Detection” (2008) PhD thesis, University of Southampton, United Kingdom
[17] P. Ekman, W. V. Friesen, and J. C. Hager, Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto, CA, USA: Consulting Psychologists Press, 2002.
[18] M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, “Measuring facial expressions by computer image analysis,” Psychophysiology, vol. 36, no. 2, pp. 253–263, Mar. 1999.
[19] Y. Tian, T. Kanade, and J. F. Cohn, “Recognizing action unites for facial expression analysis,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 2, pp. 97–115, Feb. 2001.
[20] M. S. Bartlett et al., “Recognizing facial expression: Machine learning and application to spontaneous behavior,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2. Jun. 2005, pp. 568–573.
[21] M. S. Bartlett et al., “Fully automatic facial action recognition in sponta- neous behavior,” in Proc. 7th Int. Conf. Autom. Face Gesture Recognit., Southampton, U.K., 2006, pp. 223–230.
[22] J. F. Cohn, “Foundations of human computing: Facial expression and emotion,” in Proc. Int. Conf. Multimodal Interfaces, 2006, pp. 223–238.
[23] J. F. Cohn, L. Reed, Z. Ambadar, J. Xiao, and T. Moriyama, “Automatic analysis and recognition of brow actions and head motion in spontaneous facial behavior,” in Proc. IEEE Int. Conf. Syst., Man, Cybern., 2004, pp. 610–616.

[24] B. Jiang, M. F. Valstar, and M. Pantic, "Action unit detection using sparse appearance descriptors in space-time," in Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. Workshops, Santa Barbara, CA, USA, 2011, pp. 314–321.

[25] M. F. Valstar, M. Pantic, Z. Ambadar, and J. F. Cohn, "Foundations of human computing: Facial expression and emotion," in Proc. Int. Conf. Multimodal Interfaces, 2006, pp. 162–170.

[26] G. Guo and C. R. Dyer, "Learning from examples in the small sample case—Face expression recognition," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 35, no. 3, pp. 477–488, Jun. 2005.

[27] M. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 12, pp. 1357–1362, Dec. 1999.

[28] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, Jul. 2002.

[29] C. Shan, "Smile detection by boosting pixel differences," IEEE Trans. Image Process., vol. 21, no. 1, pp. 431–436, Jan. 2012.

[30] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," Image Vis. Comput., vol. 27, no. 6, pp. 803–816, May 2009.

[31] P. Yang, Q. Liu, and D. N. Metaxas, "Exploring facial expressions with compositional features," in Proc. Int. Conf. Comput. Vis. Pattern Recognit., San Francisco, CA, USA, 2010, pp. 2638–2644.

[32] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual and spontaneous expressions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 1, pp. 39–58, Jan. 2009.

[33] Y. Chang, C. Hu, R. Feris, and M. Turk, "Manifold based analysis of facial expression," Image Vis. Comput., vol. 24, no. 6, pp. 605–614, Jun. 2006.

[34] N. Sebe et al., "Authentic facial expression analysis," in Proc. 6th IEEE Int. Conf. Autom. Face Gesture Recognit., 2004, pp. 517–522.

[35] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expres- sion database for facial behavior research," in Proc. Int. Conf. Autom. Face Gesture Recognit., Southampton, U.K., 2006, pp. 211–216.

[36] Komal D. Khawale* , D. R. Dhotre " To Recognize Human Emotions Based on Facial Expression Recognition : A Literature Survey " International Journal of Scientific Research in Computer Science, Engineering and Information Technology 2017 IJSRCSEIT | Volume 2 | Issue 1

[37] Muchiri , Ismail Ateya , Gregory Wanyembi "Human Gait Indicators of Carrying a Concealed Firearm : A Skeletal Tracking and Data Mining Approach Henry " International Journal of Scientific Research in Computer Science, Engineering and Information Technology 2018

Authors Profile

Mr. Pratik Nimbal is pursuing his Masters in Computer science and Technology (2017-2019) at Reva University, Bengaluru. He has published one paper for Face detection and face recogntion under UGC approved journal.

Dr. Gopal Kirshna Shyam received BE and M.Tech and Ph.D in Computer science and engineering from VTU, Belagavi. He has handled several subjects for UG/PG Students like Algorithms, Computer Networks, Web programming, Advanced Computer architecture, Information security, Computer Concepts and C Programming. His research interest includes Cloud Computing, Grid computing, High performance computing etc. He has published about 10 papers in highly reputed National/International Conferences like IEEE, Elsevier etc. and 5 papers in Journals with high impact factor like Elsevier Journal on Network and Computer Applications and International Journal of Cloud computing (INDERSCIENCE). His research articles on Cloud computing co-authored by Dr. Sunilkumar S. Manvi have been cited by several researchers. He is a lifetime member of CSI and is actively involved in motivating students/faculties to join CSI/IEEE/ACM societies.