3rd International Conference and Workshops on Recent Advances and Innovations in Engineering, 22-25 November 2018
(IEEE Conference Record # 43534)

Emotion Detection and Characterization using Facial Features

Charvi Jain
Computer Science and Engineering,Indian Institute of InformationTechnology
H.P, India
jaindirect@gmail.com

Kshitij Sawant,
Electronics and Communication Engineering Manipal University
Jaipur, India
kshitijsawant@gmail.com

Mohammed Rehman
Electronics and Communication Engineering Manipal University
Jaipur, India
rahmanrcks@gmail.com

Rajesh Kumar
Dept. of Electrical Engineering Malaviya National Institute of Technology
Jaipur, India
rkumar.ee@gmail.com

***Abstract*—The human face has peculiar and specific characteristics, therefore it becomes difficult in understanding and identifying the facial expressions. It is easy to identify the facial expression of particular person in any image sequence. If we look to automated recognition system, however, the systems available are quite inadequate and incapable ofaccurately identify emotions. The area of facial expression identification has many important applications. It is an interactive tool between humans and computers. The user, without using the hand can go-ahead with the facial expressions. Presently, the research on facial expression are on the factors i.e. sad, happy, disgust, surprise, fear and angry. This paper aims to detect faces from any given image, extract facial features (eyes and lips) and classify them into 6 emotions (happy, fear, anger, disgust, neutral, sadness). The training data is passed through a series of filters and processes and is eventually characterized through a Support Vector Machine(SVM), refined using Grid Search. The testing data then tests the data and their labels and gives the accuracy of classification of the testing data in a classification report. Various approaches, including passing the training images through Gabor filter, or transforming images using Histogram of Oriented Gradients(HOG) and Discrete Wavelet Transform(DWT) for better classification of data are implemented. The best result achieved so far is by passing the training images through Histogram of Oriented Gradients(HOG), followed by characterization by SVM, which gives an average precision of 85%.**

Keywords—Characterization,FacialExpression, Emotions, Cascade, Classification,SVM, Kernel, Grid Search, Wavelet, HOG, Precision

I. Introduction

The key is to understand the human behavior and how it reacts or interact with the environment. Computer interface provides technology which analyses human and computer interaction. Facial emotions convey the intention of a person. The emotion communicates the state of the person such as joy, sadness, anger. The human communication has one-third part of verbal communication and two third part of nonverbal communication[1]. Moreover, the facial expressions are an important means of interpersonal communication. Therefore, the facial expression is a key means for the detection of emotions. The non-verbal interaction among humans is through facial expression. The reason behind this interaction is that humans can identify the emotion in an efficient and prompt manner. Thus, there exists a demand to develop a machine, which can recognize the human emotion.

The objective of the work is to evaluate the performance using different models and their combinations. The models include Support Vector Machine, Linear Discriminant Analysis, Principal Component Analysis, Fisherface classifier, Gabor filters, Discrete Wavelet Transform, Histogram of Gradients. The methodologies were defined for the respective models and the Cohn-Kanade dataset is used as a stimulus for evaluation of the models [2]. This is followed by a comparative study of the said models and results. The usual way for doing the evaluation is for complete facial expression but the main focus was to reduce the features to eyes and lips only.

II. Face Detection, Extraction and Classification

In this section, a step-by-step approach towards the various processes taking place for fulfilling this work has been described. The final prediction of emotion is preceded by multiple processes.

The first step will be to determine the face of the person in the given input image, which is then succeeded by identifying the features (eyes and mouth). These features are then passed through their respective filters and transformation, if that is a part of the decided method. The outputs are then sent to the classifiers to get classified according to the trained data. This gives us the output emotion predicted by the system.

These processes will be explained in much more detail in the next sub-sections and will give a holistic view of how the system as a whole works with various comparative methods.

A. Face Detection and Feature Extraction

Face detection is regarded as one of the most complex problems in computer vision, due to the large variations caused due to changes in lighting, facial appearance and expressions.

Let's solve all the stages step by step. For face detection Viola Jones Algorithm is used. Though it was proposed in 2001 it is one of the simple and easiest method for face detection giving high accuracy[3].

This algorithmuses Haar based feature filters. The objective of this filter is to find the face in an image given as input.In each sub window Haar features are calculated and this difference is compared with the learned threshold that separates objects and non-objects.Haar features are weak classifiers so a large number of Haar classifiers are organized in such a way that they form a strong classifier which is called as "Classifier Cascade".

Each classifier looks at the sub window and determines if the sub window looks like a face and if it does then the next

978-1-5386-4525-3/18/$31.00 ©2018 IEEE

classifier is applied. If all the classifiers give a positive answer then face is there in the sub window otherwise size of sub window is changed and whole process is repeated till the face is detect [4].

Similarly, Eyes and mouth Cascades are used to detect eyes and mouth in the sub window in which face is detected.

B. Feature Classification: Different Approaches Towards Classification of Data

Once faces have been detected and the required features have been extracted, it is now time to put them through various methods designed to simplify and classify them into 6 emotions (happy, sadness, fear, neutral, anger, disgust).

It is important to note that each and every method described below is independent of each other in terms of application. The section heading defines the combination of filter(s)/transform(s)/classifier(s) used in the respective method.This section gives a brief account of all the methods used, along with description and methodology. Methodology highlights the application of that method in the dataset being used in this work.

- *Fisherface Classifier*

Linear Discriminant Analysis is a supervised algorithm that aims for classification of the input dataset. They analyze the sub space that matches the given vectors of the same class in a single blot of the feature presentation and the different classes. Thus it improves the ratio between the class scatter to the within class scatter. The sub space presentation of a group of face images, the outcome of the basis vector resulting that spaces are defined as Fisherfaces. They are helpful when facial images have wide differences in facial expressions and illumination. Principal component Analysis is predominantly used for dimensionality reduction in facial classification, image compression, etc.

Initially, the training data is to be reduced to at least N-c dimension using Principal Component Analysis where N represents the number of images in the images in the training set and c represent the number of classes. Thereafter Linear Discriminant Analysis is applied to further reduce the projected data.The equation for finding optimum weight is as follows:

$$(W_{OPT})^T = (W_{LDA})^T (W_{PCA})^T \qquad (1)$$

Where,
$(W_{LDA})^T$= Projection representing the reduction in PCA-space
$(W_{PCA})^T$ = Further Projection representing further reduced in LDA-space.

Both of the given projection, $(W_{OPT})^T$ has combination of PCA and LDA.The precision of classification for Fisherface classifier turns out to be 0.74.

- *Support Vector Machine*

A Support Vector Machine (SVM) is a machine-learning algorithm, utilizing supervised mechanism of classification. It classifies by utilizing a separating hyperplane discriminatively. The algorithm creates an optimal hyperplane which categorizes the testing data. The classifier in two-dimensional form represents a line which consists of classes on either side. This classifier utilizes the "kernel trick". This trick uses specific mathematical formulae to project the data into feature space of higher dimensions, where a hyperplane creates the possible boundaries among possible outputs. SVM is usually used to solve classification or regression problems. A kernel takes data as input and transforms it to the required form. The product between two points is returned by kernels. Thus, high dimensional projection becomes possible with very low computational cost. The kernels used in SVM are as follows:

Linear kernel : $k\langle x, x'\rangle = x \cdot x'$ (2)

Polynomial kernel : $k\langle x, x'\rangle = (\gamma(x \cdot x') + r)^d$ (3)

Radial biased function kernel: $k\langle x, x'\rangle = exp(-\gamma\|x - x'\|^2)$(4)

Sigmoid kernel : $k\langle x, x'\rangle = tanh(\gamma(x \cdot x') + r)$ (5)

where,
x, x' = two samples shown as feature vectors in input space
$\gamma = \frac{1}{2\sigma^2}$, where σ is free parameter
d = degree of polynomial
r = free parameter

The algorithm employed in this case is classifying the dataset of faces into six basic emotions (anger, disgust, fear, happy, neutral, sadness) by projecting it into a 3D feature space using kernels and separating them using appropriate hyperplanes. Initially, only lips were taken into consideration, and data consisting of images of lips converted into pixel arrays and their respective labels of 6 different emotions were split into training and testing data. This data has been trained by the SVM classifier and Grid Search. Grid search is a sub algorithm, which checks various values of 'C' and 'gamma' in the function and give out the best combination of the variables possible. The kernel with maximum accuracy as output for the testing data is taken into consideration. Later, images of eyes were included along with those of lips, and the resulting labeled data was passed through the SVM classifier and Grid Search, and results were obtained for the combined dataset.

- *Gabor Filter + SVM*

Gabor filters have a mask i.e. an array of pixels. The pixels are then given respective values, which are in a way, used as their weights. This array is convoluted with the entire image, pixel by pixel. Gabor filters change their values according the texture of the image. They give higher values at edges and points where the texture changes
Gabor filters are used to detect changes in texture as well as edges in images [5].

The images obtained by feature extraction are initially passed through the Gabor filter to highlight edges of the facial features, as well as their texture (See Fig. 1 and Fig. 2) and then the resulting labeled data is decomposed into pixel arrays and passed through the SVM classifier (discussed previously). The parameters of Gabor filter were set to appropriate values, and various combinations of kernels provided accuracies, out of which the maximum value was taken into consideration.

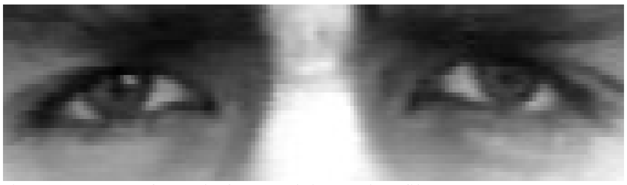

Fig 1(a). Image of eyes before applying Gabor filter

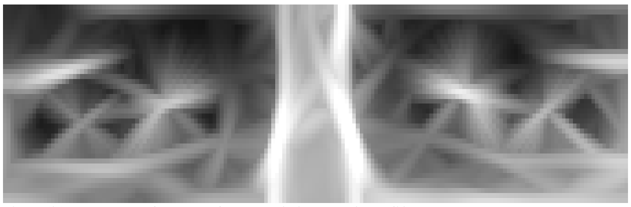

Fig 1(b). Image of eyes after applying Gabor filter

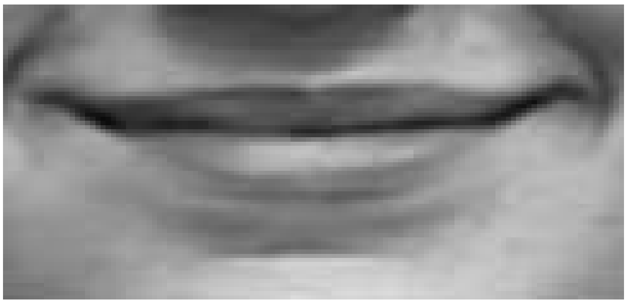

Fig 2(a). Image of mouth beforeapplying Gabor filter

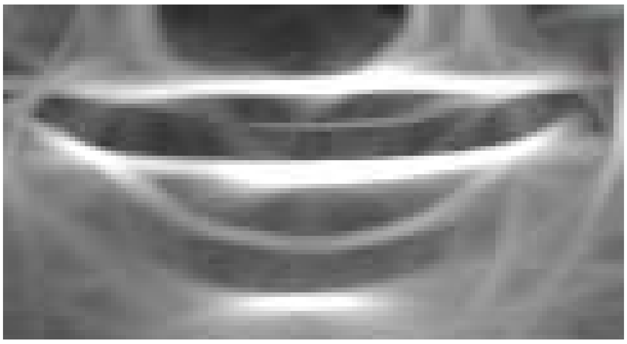

Fig 2(b). Image of mouth beforeapplying Gabor filter

- *Discrete Wavelet Transform(DWT) + SVM*

The Discreet Wavelet Transform (DWT) decomposes signals into mutually orthogonal wavelets. A wavelet is created using the scaling function. Discreet values are formed for representing the wavelet scales. The scaling properties describe the pattern of image. Other than the fact that scaling functions are mutually perpendicular to its discreet form and must stay so, DWT is an effective way to reducing noise and transforming data [6].

DWT or Discrete Wavelet Transform is a method which decomposes the image into wavelets. The features are transformed to a wavelet coefficient spectrum, consisting of certain values called signal data points. A data vector is in turn created which has the same size as our input. Scaling of wavelets occurs according to the images of eyes and mouth, and the respective data vectors are created. The data vector received through the DWT of the data is then fed to the SVM classifier, and best results are recorded.

- *Histogram Of Gradients (HOG+ SVM)*

In a HOG filter, gradients of any image are decided by distributing the image into smaller parts called cells. The description of gradients decides the orientation and magnitude of pixels in the image. These orientation bins related gradients are fixed for a cell. HOG scans the entire image and extracts the required features.

A gradient proportional to the magnitude and direction at that given pixel is decided by the cell. Image recognition and object detection algorithms get an efficient boost using this filter. It is not very useful for viewing the image, though. The feature vector produced by these algorithms produce good results, when fed into an image classification algorithm like Support Vector Machine (SVM).

The images are processed through the HOG filter to create output feature vectors of every image according to its magnitude and gradient (See Fig. 3 and Fig. 4) and are then compiled into a single dataset. This dataset describes the magnitude and vector of every pixel and is passed through SVM for classification.

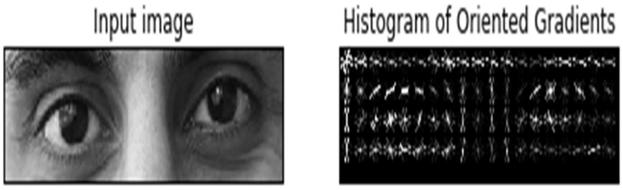

Fig 3. Image of eyes before and after applying HOG filter (left to right)

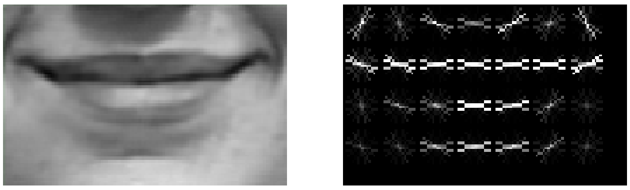

Fig 4. Image of mouth before and after applying HOG filter (left to right)

- *DWT + HOG + SVM*

The images saved after going through the Discreet Wavelet Transformation(DWT) are processed through the HOG filter to create output feature vectors of every image (See Fig. 5 and Fig. 6) and are then compiled into a single dataset. This dataset describes the magnitude and vector of every pixel and is passed through SVM for classification.

Discreet Wavelet Transform aims to convert the given image into varying scaling factors of a wave, which is further transformed into its HOG form, thus aiming to reveal features of the given face in much more clarity. The classifier receives information in form of gradients given by the HOG transformation.

TABLE 1. RESULTS OF VARIOUS CLASSIICATION METHODS CORRESPONDING TO SVM COMBINED WITH OTHER FILTERS

METHODS	Anger	Disgust	Fear	Happy	Neutral	Sad	Avg.
HOG + SVM	0.80	0.80	0.40	0.94	0.92	0.75	0.85
Gabor filter + SVM	0.88	0.67	0.75	0.90	0.74	1.00	0.81
SVM	0.86	0.54	0.80	0.94	0.69	1.00	0.80
DWT + SVM	1.00	0.50	0.50	0.78	0.71	0.33	0.72
DWT + HOG + SVM	1.00	0.36	0.50	0.87	0.65	0.29	0.70

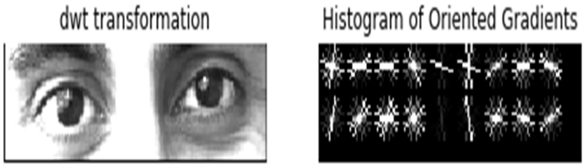

Fig 5. Image of eyes before and after applying HOG filter with DWT (left to right)

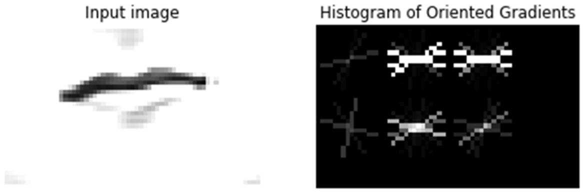

Fig 6. Image of mouth BEFORE AND AFTER APPLYING HOG filter with DWT (left to right)

III. DATASET

The Extended Cohn Kanade Dataset (CK+) has 593 numbers of image sequences (327 image sequence contains the emotion labels) from 8 facial expressions namely neutral, sadness, fear, happiness, surprise, anger, disgust, contempt. The emotions considered in this dataset are 6 out of the 8, namely neutral, sadness, fear, happiness, anger and disgust. The image sequences begin with a neutral expression and end towards a peak face expression. Each frame size has a resolution of 640 x 490 and are usually grey. To begin with, data is classified into two folders, one has collection of images and the other containing the text file [7].

IV. EXPERIMENTAL SETUP

The experiment utilizes the Cohn-Kanade dataset by extracting the extreme image showing the required expression,amongst a series of images of any individual from neutral to the labelled expression. These images are then fed to their respective filters, which finally convert the image to a CSV file consisting of data representing that image, which is fed to the classifier to predict results. Real time images are taken on mobile phones, and then processed through the system accordingly.

V.RESULTS AND COMPARATIVE STUDY

The classification of emotions carried out in previous sectionsinvolves multiple approaches towards characterizing the given dataset, that too with a combination of multiple methods(See Table 1).

The Fischerface classifier uses Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), both of which contribute to its accuracy. PCA aims to reduce dimensionality, such that variables in the dataset are reduced to a minimum. The new set of variables, called principal components make the classification much simpler in terms of space complexity. It gives a precision output of 0.74, which is fairly good, but still not enough.

Gabor filter recognizes texture in any given image and creates frequency and orientation components. This gives Gabor filter an edge over other filters, as difference in texture could be a very efficient way to differentiate mouth and eyes from the rest of the skin. These filters also decompose in multiple dimensions in space. This might also be the reason why SVM works best with this filter, as projecting in higher spatial dimensions is a property common to both the filter and the classifier in this case, thus giving a precision of 0.81.

The most precise classification of 0.85 occurs by using a combination of Histogram of Gradients (HOG) filter followed by classification by SVM. This combination has been used frequently in the past due to its high precision rate, and due to valid reason. The method by which HOG partitions the image into boxes, so as to provide features of every box available, makes the data much more precise, and reduces noise from the image. The classifier thus gets an image which not only has less noise, but also has a strong sense of direction. Gabor filters are not as powerful in cases where the texture is more complex, such as in the face of disgust, but in such cases, HOG filters have the upper hand, due to simplicity of data. Simply applying SVM also gives a precision of 0.80, which is due to the ability of a simple SVM algorithm to project the pixel densities of trained images into higher dimensions.

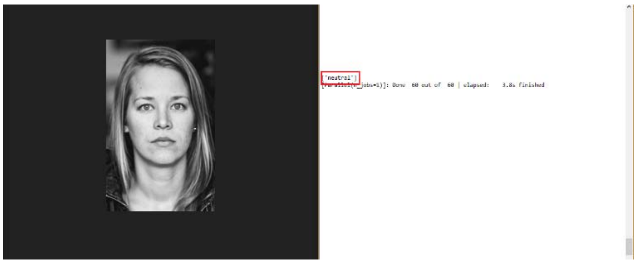

Fig 7. Input face image and predicted result “neutral”(Classification method = HOG + SVM) [8]

Code Output (Right half of image):

```
[‘neutral’]
[Parallel(n_jobs=1)]: Done 60 out of 60 | elapsed: 3.8s finished
```

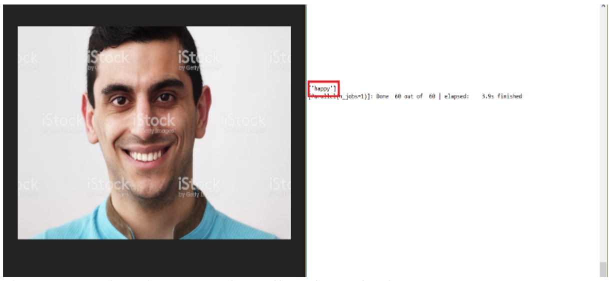

Fig 8. Input face image and predicted result “happy” (Classification method = HOG + SVM) [9]

Code Output (Right half of image):

```
[‘happy’]
[Parallel(n_jobs=1)]: Done 60 out of 60 | elapsed: 3.9s finished
```

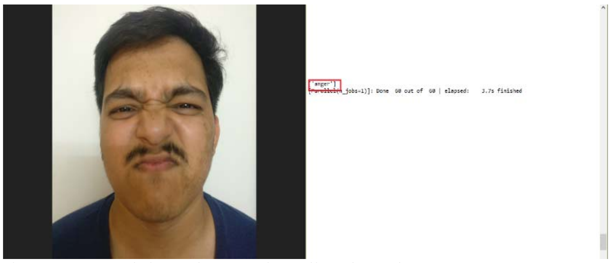

Fig 9. Input real time photo and predicted result “angry” (Classification method = HOG + SVM)

Code Output (Right half of image):

```
[‘anger’]
[Parallel(n_jobs=1)]: Done 60 out of 60 | elapsed: 3.7s finished
```

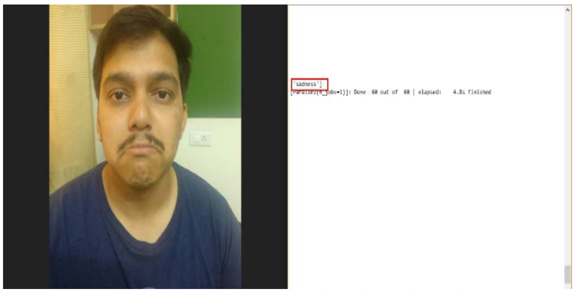

Fig 10. Input real time photo and predicted result “sadness” (Classification method = HOG + SVM)

Code Output (Right half of image):

```
[‘sadness’]
[Parallel(n_jobs=1)]: Done 60 out of 60 | elapsed: 4.0s finished
```

Discrete Wavelet Transform (DWT) creates sub-signals in the horizontal, vertical and diagonal direction, which are then analyzed to gain a general form of the image taken.This transformation is very useful in detecting patterns, or abnormalities among regular trends. In face detection however, DWT falls back as compared to Gabor filter and HOG. This happens due to the variation in every image that has to be trained. Change of skin tone, differing features in the greyscale image of a face can tend to confuse the Wavelet specifications, thus giving a precision of 0.72.

Finally, passing the images through a DWT filter followed by an HOG filter and then classifying by SVM gives us a precision of 0.70. This is bound to occur, as transforming the image to DWT form, as seen before, has reduced the precision to 0.72. The Wavelet features obtained by the output of DWT do not provide any sense of direction or magnitude, which is exactly what the HOG filter is supposed to receive and analyze. This leads to unnecessary information being passed to the HOG filter, which the classifier classifies with a precision of 0.70. Also, horizontal details of mouth might appear very similar, thus resulting in decrease of precision.

When we overview all the used filters above, it is eminent that the main contributors of classification are the directionality and magnitude of a facial image, followed by the texture differences in that particular image. Wavelet transformation seems to provide satisfactory results, whereas the combination of HOG followed by SVM classifier does not disappoint.

VI. CONCLUSION

Emotions are an integral method of expressing our judgement and decisions in daily life, and this work aims to recognize and detect exactly these emotions.This work is capable of recognizing 6 integral emotions – Happy, Sad, Anger, Fear, Neutral and Disgust; with the help of the Support Vector Machine algorithm. This image primarily uses only 2 crucial features of the face, namely eyes and mouth, to detect an emotion within a face. The Viola Jones algorithm is utilized in order to detect face and features of the individual in the input photo. These features include eyes and mouth.

The predictions occurred by simply applying SVM are not as accurate as when an HOG filter is applied before classification. This happens due to lack of distinguishable features without the HOG transform. The advantage of this system is that it needs only 2 features for detecting the emotion of a complete face, which satisfyingly decreases amount of storage data necessary for testing and for future applications.

ACKNOWLEDGMENT

The authors would like to thank all the members ofRAMAN Lab,MNIT especially Ms. Vishu Gupta for providing excellent guidance. Theyare grateful to them for this opportunity and support.

REFERENCES

[1] Byoung Chul Ko. ABrief Review of Facial EmotionRecognition Basedon Visual Information, Sensors 2018, 18, 401.

[2] Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG’00), Grenoble, France, 46-53.

[3] Paul Viola , Michael Jones, Rapid object detection using a boosted cascade of simple features, Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001.

[4] Prof. Neelum Dave, NarendraPatil, RohitPawar, Digambarpople, Emotion Detection Using Face Recognition ,IJESC Volume 7 Issue No.4.

[5] T. Ahsan, T. Jabid, and U.-P. Chong,Facial expression recognition using local transitional pattern on Gabor Filtered facial image,. IETE Technical Review, 30(1):47{52, 2013}.

[6] H.K. Meena, K.K. Sharma and S.D. Joshi, Improved facial expression recognition using graph signal processing, Electronics Letters (Volume: 53, Issue: 11, 5 25 2017).

[7] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101.

[8] Image taken from “http:/ /i1.wp.com/ detourphotography.ca /wp-content/ uploads/2015/11/DSC_8822.jpg ?resize = 236 % 2C355”. Accessed on 21 Aug.2018.

[9] Image taken from “https://www.dreamstime.com/close-up-happy-middle-eastern-man-s-face-white-background-image 111174542”. Accessed on 21 Aug. 2018.