ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Neural Network Based Expression Recognition Using Active Facial Patches

Swathy A S[1], Suby Varghese[2]

PG Scholar, Dept. of ECE, TKM Institute of Technology, Kollam, Kerala, India[1]

Asst. Professor, Dept. of ECE, TKM Institute of Technology, Kollam, Kerala, India[2]

ABSTRACT: Facial expression, being a fundamental mode of communicating human emotions, finds its applications in human-computer interaction (HCI), health-care, surveillance, driver safety, deceit detection etc. For effective facial expression recognition, the extraction of discriminative features from facial patches is necessary. So, here a technique based on the appearance features of selected facial patches is used for expression recognition. A few prominent facial patches, depending on the position of facial landmarks, are extracted which are active during emotion elicitation and the features from these active patches are given to artificial neural network (ANN) to discriminate different expressions. The method is found to perform well consistently on JAFFE database using MATLAB software.

KEYWORDS: facial landmark detection, feature extraction, facial patches, ANN

I. INTRODUCTION

Facial expression is a visible manifestation of affective state, cognitive activity, intention, personality, and psychopathology of a person; it not only expresses our emotions but also provides important communicative cues during social interaction. Reported by psychologists, facial expression constitutes 55% of the effect of a communicated message while language and voice constitute 7 and 38%, respectively. So it is obvious that automatic recognition of facial expression can improve human-computer interaction (HCI) or even social interaction. Facial expression recognition (FER) can be useful in many areas, for research and application. Studying how humans recognize emotions and use them to communicate information is important topic in anthropology. The emotion automatically estimated by a computer is considered to be more objective than those labeled by people and it can be used in clinical psychology, psychiatry, and neurology. Furthermore, expression recognition can be embedded into a face recognition system to improve its robustness.

Effective expression analysis hugely depends upon the accurate representation of facial features. Facial Action Coding System (FACS) represents face by measuring all visually observable facial movements in terms of Action Units (AUs) and associates them with the facial expressions. Accurate detection of AUs depends upon proper identification and tracking of different facial muscles irrespective of pose, face shape, illumination, and imageresolution. But the detection of all facial fiducial points is even more challenging than expression recognition itself. Therefore, most of the existing algorithms are based on geometric and appearance based features. The models based on geometric features track the shape and size of the face and facial components such as eyes, lip corners, eyebrows etc., and categorize the expressions based on relative position of these facial components. However, these methods usually require very accurate and reliable detection as well as tracking of the facial landmarks which are difficult to achieve in many practical situations. Moreover, the distance between facial landmarks vary from person to person, thereby making the person independent expression recognition system less reliable. Changes in facial expressions involve contraction and expansion of facial muscles which alters the position of facial landmarks [1]. Along with the facial muscles, the texture of the area also changes. This method attempts to understand the contribution of different facial areas toward automatic expression recognition. In other words, the work explores the facial patches which generate discriminative features to separate two expressions effectively.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

II. RELATED WORK

An automatic classification of facial expressions consists of two stages: feature extraction and feature classification. The feature extraction is a key importance to the whole classification process. If inadequate features are used, even the best classifier could fail to achieve accurate recognition. The two most common approaches to the facial feature extraction are the geometric feature-based methods and the appearance - based methods. Geometric features present the shape and locations of facial components (including mouth, eyes, eyebrows and nose). The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. The appearance features present the appearance (skin texture) changes of the face, such as wrinkles and furrows. The appearance features can be extracted on either the whole face or specific regions in a face image. Here, different feature extraction and classification methods used in effective expression recognition are included.

A. *FEATURE EXTRACTION METHODS*

The feature extraction process converts pixel data into a higher-level representation of shape, motion, color, texture, and spatial configuration of the face or its components. The extracted representation is used for subsequent classification. Feature extraction generally reduces the dimensionality of the input space. The reduction procedure should retain essential information possessing high discrimination power and high stability.

i. **Gabor wavelet filters:** A Gabor filter is a function obtained by modulating the amplitude of a sinusoid with a Gaussian function. They can be applied to images to extract features aligned at particular angles (orientations). The most important parameters of a Gabor filter are its orientation and frequency. Certain features that share similar orientation or frequency can be selected and used to differentiate between different facial expressions depicted in images [2]. A two-dimensional Gabor filter is expressed as Gaussian modulated sinusoid in the spatial domain, and as a shifted Gaussian in the frequency domain.

ii. **Canny edge detector:** The Canny edge detector is an edge detection operator that uses a multi-stage algorithm to detect a wide range of edges in images. It is a technique to extract useful structural information from different vision objects and dramatically reduce the amount of data to be processed. It has been widely applied in various computer vision systems [3]. The Process of Canny edge detection algorithm can be broken down to 5 different steps.

- Apply Gaussian filter to smooth the image in order to remove the noise
- Find the intensity gradients of the image
- Apply non-maximum suppression to get rid of spurious response to edge detection
- Apply double threshold to determine potential edges
- Track edge by hysteresis: Finalize the detection of edges by suppressing all the other edges that are weak and not connected to strong edges.

iii. **Histogram of oriented gradients:** The histogram of oriented gradients (HOG) is a feature descriptor used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image. The image is divided into small connected regions called cells, and for the pixels within each cell, a histogram of gradient directions is compiled. The descriptor is the concatenation of these histograms. For improved accuracy, the local histograms can be contrast-normalized by calculating a measure of the intensity across a larger region of the image, called a block, and then using this value to normalize all cells within the block.

B. *CLASSIFICATION METHODS*

The goal of classification is to classify an instance to a class based on the value of several attributes. Many approaches to classification attempt to explicitly construct a function from the joint set of values of the attributes to class labels

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

i.**Bayesian classifier:**In machine learning, Bayesian classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. The idea behind this classifier is that, if an agent knows the class, it can predict the values of the other features. If it does not know the class, Bayes' rule can be used to predict the class given the feature values. In a Bayesian classifier, the learning agent builds a probabilistic model of the features and uses that model to predict the classification of a new example. In this approach, we approximate the joint probability distribution of the class and the attributes : $P_r(C,A_1,\ldots,A_k)$, where C is a random variable describing the class, and $A_1,\ldots,A_k$ are random variables describing the attributes. Thus learning in Bayesian classification amounts to estimation of this joint probability distribution. After we construct such an estimate, we classify new instances by examining conditional probability of C given the particular attribute values, and returning the class that is most probable.

ii**. Support Vector Machine (SVM) Classifier:** SVM was used for classification of extracted features into different expression categories. SVM is a popular machine learning algorithm which maps the feature vector to a different plane, usually to a higher dimensional plane, by a non-linear mapping, and finds a linear decision hyper plane for classification of two classes [4]. Consider a training sample (x_i ,y_i) where x_i is the input pattern for i^{th} example and y_i is the desired/target output. Assume that the class represented y_i = +1 and class represented by y_i = -1 are linearly separable. Then the hyper plane H can be defined as

$$\mathbf{x}_i \bullet \mathbf{w} + b \geq +1 \text{ when } y_i = +1 \quad (1)$$

$$\mathbf{x}_i \bullet \mathbf{w} + b \leq -1 \text{ when } y_i = -1 \quad (2)$$

III. METHODOLOGY

The overview of the proposed method is shown in Fig 1. The accurate facial landmark detection and extraction of appearance features from active face regions improve the performance of expression recognition. Therefore, the first step is to localize the face followed by detection of the landmarks. A learning-free approach is used in which the eyes and nose are detected in the face image and a coarse region of interest (ROI) is marked around each. The lip and eyebrow corners are detected from respective ROIs [1]. Locations of active patches are defined with respect to the location of landmarks. Then the features from active facial patches are extracted and classified into different expressions using a multiclass classifier.

The training phase includes pre-processing, selection of facial patches, extraction of appearance features and learning of the multi-class classifiers. In an unseen image, the process first detects the facial landmarks, then extracts the features from the facial patches, and finally classifies the expressions.

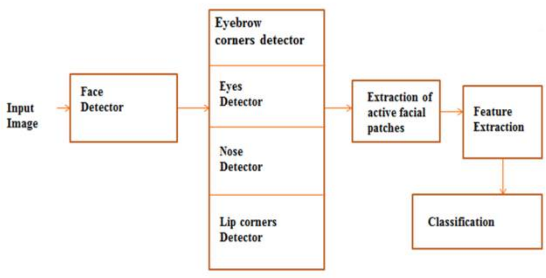

Fig 1 : Overview of the system

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

A. FACIAL LANDMARK DETECTION

The facial patches which are active during different facial expressions are positioned below the eyes, in between the eyebrows, around the nose and mouth corners. To extract these patches from face image, it is need to locate the facial components first followed by the extraction of the patches around these organs. Here, a robust, learning-free, lightweight generic face model fitting method for localization of the facial organs are used. Using local gradient analysis, this method finds the facial features and adjusts the deformable 3D face model so that its projection on image will match the facial feature points. The active facial patches with respect to the position of eyes, eyebrows, nose, and lip corners are extracted using the geometrical statistics of the face. Fig 2 shows the steps involved in automated facial landmark detection and patch extraction.

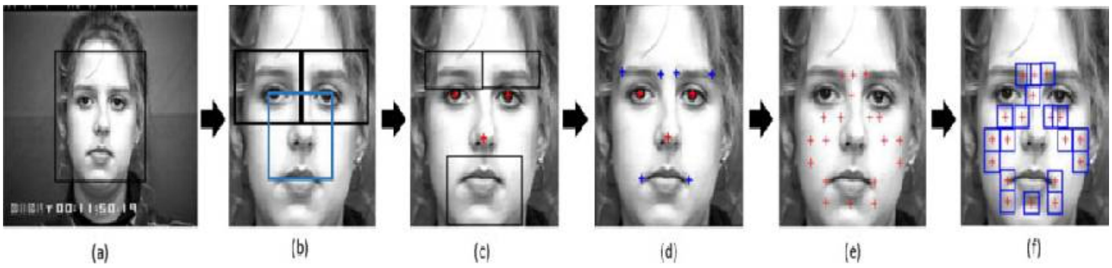

Fig 2 : Framework for automated facial landmark detection

i. Preprocessing: A low pass filtering was performed using a 3x3 Gaussian mask to remove noise from the facial images followed by face detection for face localization. Viola-Jones technique of Haar-like features with Adaboost learning is used for face detection [5]. It has lower computational complexity and was sufficiently accurate for detection of near-frontal and near-upright face images. Using integral image calculation, it can detect face regardless of scale and location in real time. The localized face was extracted and scaled to bring it to a common resolution. This made the algorithm shift invariant, i.e., insensitive to the location of the face on image.

ii. Eye and Nose Localization: To reduce the computational complexity as well as the falsedetection rate, the coarse region of interests (ROI) for eyes and nose were selected using geometrical positions of face. Both the eyes were detected separately using Haar classifiers trained for each eye. The Haar classifier returns the vertices of the rectangular area of detected eyes. The eye centers are computed as the mean of these coordinates. Similarly, nose position was also detected using Haar cascades. In case the eyes or nose was not detected using Haar classifiers, the system relies on the landmark coordinates detected by anthropometric statistics of face. The position of eyes was used for up-right face alignment as the positions of eyes do not change with facial expressions

iii.Lip Corner Detection: The ROIs for lips and eyebrows were selected as a function of face width positioned with respect to the facial organs. The ROI for mouth was extracted using the position of nose as reference. The upper lip always produces a distinct edge which can be detected using a horizontal edge detector. Sobel edge detector was used for this purpose. In images with different expressions, a lot of edges were obtained which was further threshold by using Otsu method. In this process, a binary image was obtained containing many connected regions. Using connected component analysis, the spurious components having an area less than a threshold were removed. Further, morphological dilation operation was carried out on the resulting binary image. Finally, the connected component with largest area which was just below the nose region was selected as upper lip region.

iv.Eyebrow Corner Detection: With the knowledge of positions of eyes, the coarse ROIs of eyebrows were selected. The eyebrows were detected following the same steps as that of upper lip detection. However, performing an adaptive threshold operation before applying horizontal sobel operator improved the accuracy of eyebrow corner localization. The use of horizontal edge detector reduced the false detection of eyebrow positions due to partial occlusion by hair.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

v. Extraction of active facial patches: During an expression, the local patches were extracted from the face image depending upon the position of active facial muscles. The active patches does not have very fixed position on the face image. Rather, their location depends upon the positions of facial landmarks [6]. The size of all facial patches was kept equal and was approximately one-ninth of the width of the face. Here onwards, we will refer the patches by the numbers assigned to it. As shown in Fig 3, P_1, P_4, P_{18}, and P_{19} were directly extracted from the positions of lip corners and inner eyebrows respectively. P_{16} was at the center of both the eyes; and P_{17} was the patch above P_{16}. P_3 and P_6 were located in the midway of eye and nose. P_{14} and P_{15} were located just below eyes. P_2, P_7, and P_8 were clubbed together and located at one side of nose position. P_9 was located just below P_1. In a similar fashion P_5, P_{11}, P_{12}, and P_{13} were located. P_{10} was located at the center of position of P_9 and P_{11}.

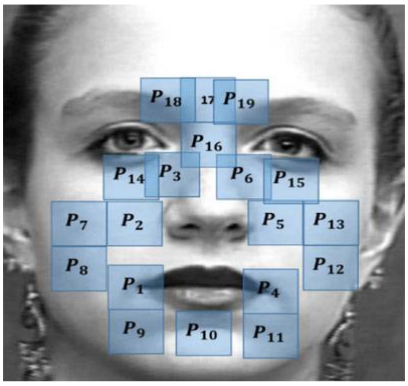

Fig 3 : Position of facial patches

B. FEATURE EXTRACTION AND CLASSIFICATION

i. **Feature extraction:**Local Binary Patterns were introduced as an effective texture descriptors. Input image is transformed into LBP representation by sliding window technique where value of each pixel in the neighborhood is thresholded with value of central pixel [7]. Central pixel is encoded with LBP code (binary or decimal) in corresponding LBP image pixel (Fig 4). Binary codes are so called 'micro-textons' that represent texture primitives such as curved edges, flat or convex areas. Basic version of LBP uses 3x3 sliding window to code the texture. Recently, the operator has been extended to different sizes and shapes (circular neighborhood). The size of the neighborhood directly influences the range of code values. Having operator of size P and radius R, the range of possible codes are from 0 to 2^P. The image texture is described by a 2^P bin histogram of corresponding LBP image.

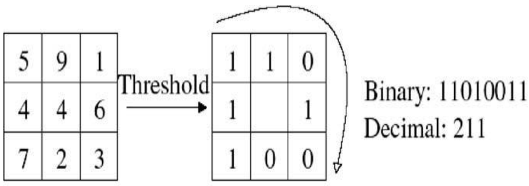

Fig 4 : LBP encoding

ii. **Multiclass classification using ANN**: Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the network function is determined largely by the connections between elements. We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements. Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output. The network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Typically many such input/target pairs are used, in this supervised learning, to train a network.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

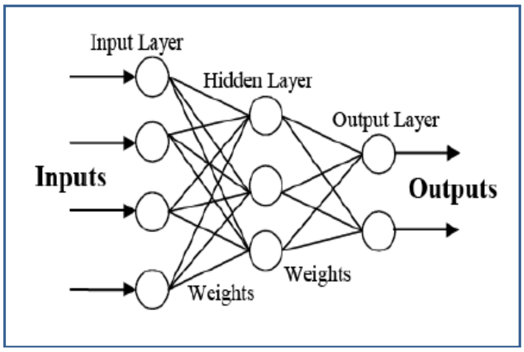

Fig 5 : Architecture of artificial neural network

An ANN consists of a number of interconnected artificial processing neurons called nodes, connected together in layers forming a network. A typical ANN is schematically illustrated in Fig 5. Here, each node in a layer provides a threshold of single value by summing up their input value pi with the corresponding weight value wi . Then the neuron's net input value n is formed by adding up this weighted value (sum), with the bias term b. The bias is added to shift the sum relative to the origin (Fig 6).

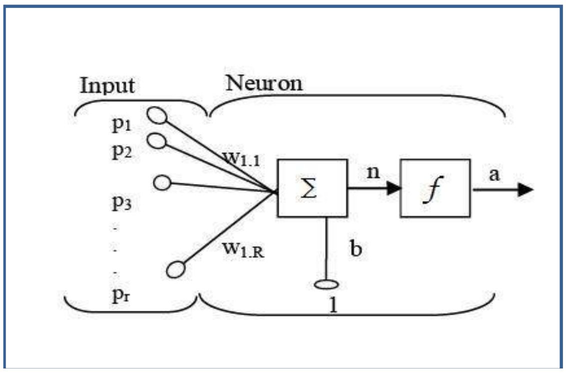

Fig 6 : A neuron

IV. RESULTS AND DISCUSSIONS

The method was evaluated by using Japanese Female Facial Expressions (JAFFE) database. As discussed earlier, face detection was carried out on all images followed by scaling to bring the face to a common resolution. Facial landmarks were detected and active facial patches were extracted from each face image. During training stage, an ANN was trained between each pair of expressions. Here the training data were the concatenated LBP histogram features extracted from the active facial patches containing discriminative characteristics between the given pair of expression classes.

Viola-Jones technique of Haar-like features with Adaboost learning is used for face detection. The method returns a bounding box around the face region. The same method can be used to identify different region of interests (ROIs) by changing the model name in the algorithm, ie ., lips ,eyebrows, nose and eyes were detected. Using this coarse regions , the different facial landmarks were detected. Both the eyes were detected separately using Haar classifiers trained for each eye. The Haar classifier returns the vertices of the rectangular area of detected eyes. The eye centers arecomputed as the mean of these coordinates. Similarly, nose position was also detected using Haar cascades. The lip corners and eyebrow corners were found using the same algorithm shown in Fig 7.

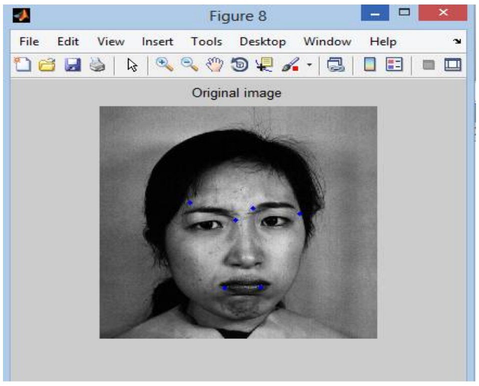

Fig 7 : Facial landmark detection

During an expression, the local patches were extracted from the face image depending upon the position of active facial muscles.Some facial patches are common during elicitation of all basicexpressions and some are confined to a single expression. These active patches are positionedbelow the eyes, in between the eyebrows, around the noseand mouth corners. The active patches do not have very fixed position on the face image. Rather, their location depends upon the positions of above detected facial landmarks. The size of all facial patches was kept equal and was approximately one-ninth of the width of the face. The active facial patches extracted based on the location of these facial landmarks are shown in Fig 8.

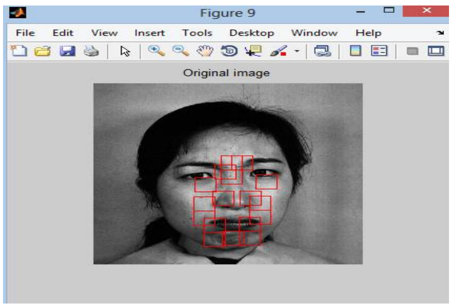

Fig 8 : Extraction of active facial patches

From these active facial patches the features were extracted using local binary pattern.The LBP features of all the active facial patches are concatenated to provide one feature vector. Similarly the feature vectors for all the input images are computed and are given to the input of an ANN for training. During testing, the method extracts the LBP features from active facial patches and classifies into different expressions shown in Fig 9.

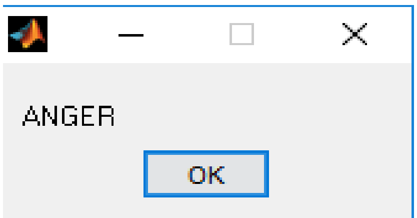

Fig 9 : Output of ANN

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

V. CONCLUSION

The simulation results showed that the described algorithm is an efficient method for the classification of six universal expressions. It investigates the relevance of different facial patches in the recognition of different facial expressions. All major active regions on face are extracted which are responsible for the face deformation during an expression. The position and size of these active regions are predefined. From these active patches the features are extracted and are given to a multiclass classifier to discriminate expressions depicted in images.

REFERENCES

1. S L Happy and Aurobinda Routray, "Automatic Facial Expression Recognition Using Features of Salient Facial Patches" , IEEE Trans. Affective Comput., vol. 6, no. 1, January-march 2015.
2. 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..
3. Y. Tian, T. Kanade, and J. F. Cohn, "Recognizing lower face action units for facial expression analysis," in Proc. IEEE Int. Conf. Autom. Face Gesture Recog., 2000, pp. 484–490.
4. C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 415–425, Mar. 2002.
5. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2001,pp.511–518.
6. [10] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. N. Metaxas, "Learning active facial patches for expression analysis," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2012, pp. 2562–2569
7. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution grayscale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, Jul. 2002.
8. R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," IEEE Trans.Affective Comput., vol. 1, no. 1, pp. 18–37, Jan. 2010.
9. L. Unzueta, W. Pimenta, J. Goenetxea, L. P. Santos, andF. Dornaika, "Efficient generic face model fitting to images and videos," *Image Vis. Comput.*, vol. 32, no. 5, pp. 321–334, 2014.
10. S. M. Lajevardi and Z. M. Hussain, "Automatic facial expression recognition: Feature extraction and selection," *Signal, Image Video Process.*, vol. 6, no. 1, pp. 159–169, 2012.

BIOGRAPHY

SwathyA S is a PG scholar in Electronics and Communication Department, TKM Institute of Technology, Kollam, Kerala, India. She received Bachelor of Technology (B Tech) degree in 2015 from College of Engineering Perumon, Kollam, Kerala, India.

Suby Varghese is working asAssistant Professor in Electronics and Communication Department, TKM Institute of Technology, Kollam, Kerala, India. She received Master of Engineering (ME) in Communication System at PSN College of Engineering & Technology under Anna University in 2011 and Bachelor of Technology (B Tech) degree in 2007 from Caarmel Engineering College, Perunad, Pathanamthitta, Kerala, India.