Efficient Facial Emotion Classification With Wavelet Fusion of Multi Features

K.V. Krishna Kishore[1], G.P.S.Varma[2]

1. Department of Computer Science & Engineering, Vignan University, Guntur, A.P,India.
2. Department of Information Technology, S.R.K.R. Engineering College, Bhimavaram, A.P, India

Abstract

In this paper, a new facial emotion classifier is proposed based on wavelet fusion, which combines the features extracted by Gabor wavelet and Discrete Cosine Transform (DCT). We show that combining two of the most successful methods such as Gabor wavelets and DCT gives considerably better performance than either alone: they are complementary in the sense that DCT captures global features while Gabor extracts local features. Both feature sets are high dimensional so it is beneficial to use Principle Component Analysis (PCA) and to reduce the dimensionality of data. Finally, we introduce Wavelet fusion to fuse local features of Gabor and global features of DCT. The proposed approach is evaluated on Cohn-Kanade database. In particular, we perform comparative experimental studies of independent methods with multi feature methods. We also make a detailed comparison of different fusion techniques with wavelet fusion, as well as different Neural Network classifiers. Extensive experimental results verify the effectiveness of our approach outperforms most of the state-of-the-art approaches.

Keywords*:* *Gabor wavelet, DCT, PCA, Wavelet fusion, RBF, PNN*

1. Introduction

In recent years, classification of facial emotions has become one of the most important challenges in computer vision. It is argued that the facial emotion play an important role in social interactions with other human beings. Facial emotions can contain a great deal of information and the desire to automatically extract this information has been continuously increasing. With the rapid development of computer vision and artificial intelligence, facial emotion recognition becomes the key technology of advanced human computer interaction. Emotion is common word used for what a person is feeling at a given moment joy, sad, anger, surprise, fear and disgust. According to Ekman, et al., [3] "Basic" emotions are distinct families of affective states characterized by different signals, physiology, appraisal mechanisms, and antecedent events. Ekman cites early evidence suggesting that each emotion is accompanied by distinctive physiological changes that prepare an organism to respond appropriately. Just as emotions are critical to human behavior, they are equally critical for intelligent machines, especially autonomous machines of the future that will help people in their daily activities. Robots, to be successful, will have to have emotions. The machines and products of the future may be able to sense human emotions and respond accordingly, but they will not be smart and sensible until they have both intelligence and emotions. Facial Emotion Recognition and classification nothing but recognizes the persons when they are in emotional state and classify the emotions into 6 basic emotions. In most cases of facial emotion classification, the process of feature extraction yields a definitively large number of features and subsequently a smaller sub-set of features needs to be selected according to some optimality criteria.

The best classifier in the world, human beings, also do not rely on a single modality to recognize objects if they are producing different information. Information fusion covers any area, which deals with utilizing a combination of different sources of information, either to generate one representational format, or to reach a decision. Feature-level methods combine several incoming feature sets into a single fused one then it is used with a conventional classifier, whereas decision-level ones combine several classifiers (e.g. based on distinct features) to make a stronger final classifier.

Multi-features system has an attractive research direction in biometrics. A multi-feature system usually contains two or more kinds of features extracted from the same biometric trait and merges them together for personal identification. Hence, compared to single-feature approaches, multi-features fusion approaches always have better in accuracy and stability. For multi-feature fusion methods, many works have been done at different fusion levels (feature level, matching score level and decision level) [1]. The fusion at the feature level, compared to the others, was expected to perform better in emotion recognition when the discriminability of features from

Manuscript received August 5, 2011
Manuscript revised August 20, 2011

different sets could be preserved mostly. Hence, in this study, attention is concentrated on the feature level fusion. Gabor filters [2,5,7] have been proved to be effective for emotion recognition because of its superior capability of multi-scale representation. Gabor wavelet [2,5,7] can use very better description of biological visual neuron about receptive field. According to the needs of special vision, it can adjust the spatial and frequency properties to face emotion characteristic wanted, so Gabor filter wavelet is suitable for people face analysis and treatment of emotion. The remainder of study provides details on the methods, experiments and results. Section 2, describes the proposed architecture, the feature extraction method based on the Gabor Filter, DCT and wavelet fusion. Section 3 presents the experimental results, and Section 4 contains the conclusion.

2. The Proposed Method

The overall architecture of the proposed facial emotion classification method is given in Fig. 1

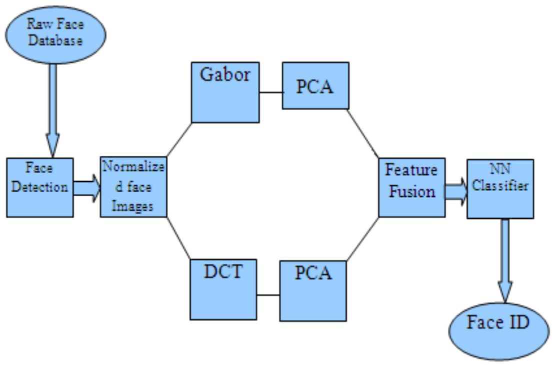

Fig1. Proposed architecture of multi future facial emotion recognition using wavelet fusion

Face detection algorithm is used to detect faces from raw image files and are cropped to eliminate background noise in the face images. Then, normalization method applied to eliminate variations in illumination of face images. In order to achieve high recognition rate, the selection of feature extraction algorithm is very crucial. In this study, the features are extracted with Gabor wavelet and DCT algorithm. The resultant features are at high dimensional space, so dimensionality reduction algorithm Principal Component Analysis (PCA) [4,8] is used to reduce the dimensions. Then the resultant features are fused with wavelet fusion and compare with existing fusion techniques.

2.1 Gabor Wavelets

At first Gabor Wavelets [2,5,7] are used to extract the features from the static images. Gabor functions provide the optimized resolution in both the spatial and frequency domains. Gabor wavelets seem to be optimal basis to extract local features for pattern recognition. The whole set of 40 Gabor kernels are generated, with magnitude at five scales and eight different orientations. The real parts of Gabor wavelets are shown in the Fig.2. The Gabor wavelets (kernels, filters) can be defined as follows

$$G_{u,v}(x,y) = \frac{(u^2+v^2)}{2\sigma^2} e^{\frac{-((u^2+v^2)(x^2+y^2))}{2\sigma^2}} (e^{i(ux+vy)} - e^{\sigma^2/2})$$

$$\text{Where } u = \left(\gamma/f^{nu}\right)\cos\left(pi * {}^{mu}/_{8}\right)$$

$$v = \left({}^{-\gamma}/f^{nu}\right)\sin\left(pi * {}^{mu}/_{8}\right)$$

$$\gamma = {}^{pi}/_{2} \text{ And } f = \sqrt{2}$$

The wavelets exhibit desirable characteristics of spatial frequency, spatial locality and orientation selectivity.

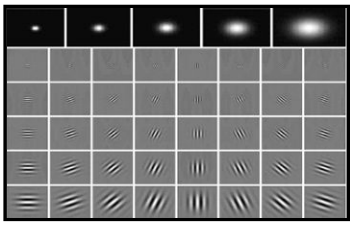

Fig 2: Real Part and Imaginary Parts of Gabor kernel

The dimension of a Gabor feature vector is so high that the computation and memory requirements are very large. Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels. Let I(x, y) be the gray level distribution of an image, the convolution of image I and a Gabor kernel $\psi_{u,v}$ is defined as follows:

$$O_{u,v} = I(z) * \psi_{u,v}(z)$$

Where z=(x, y), * denotes the convolution operator and $O_{u,v}(z)$ is the convolution result corresponding to the Gabor kernel at orientation u and scale v. Therefore the set $S=\{O_{u,v}(z): u\epsilon\{0\ldots7\}, v\epsilon\{0\ldots4\}\}$ forms the Gabor wavelet representation of the image I(z). Applying the convolution theorem to derive each $O_{u,v}(z)$ from above equation via Fast Fourier Transform(FFT):

$$\mathcal{F}\{O_{u,v}(z)\} = \mathcal{F}\{I(z)\mathcal{F}\{\psi_{u,v}(z)\}\}$$

and $O_{u,v}(z) = \mathcal{F}^{-1}\{\mathcal{F}\{I(z)\}\mathcal{F}\{\psi_{u,v}(z)\}\}$ where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier and inverse Fourier transform, respectively

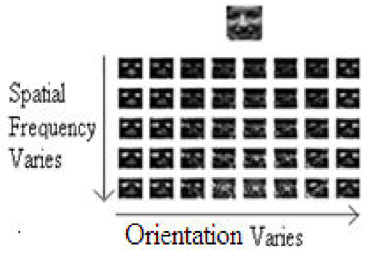

Fig 3: Variation of Scaling and rotation with Gabor kernel

2.2 Discrete Cosine Transform

Global features are extracted by using the DCT [9] and the large area illumination variations are alleviated by discarding the first few low-frequency DCT coefficients [9]. Actually, the image is broken in to 8*8 blocks of pixels, from left to right and top to bottom then DCT is applied to each block. Each block is compressed through quantization. The array of compressed blocks that constitute the image is stored in a drastically reduced amount of space. The DCT equation computes the i, jth entry of the image.

$$D(i,j) = \frac{1}{\sqrt{2N}} C(i)C(j) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} p(x,y) \cos\left[\frac{(2x+1)i\pi}{2N}\right] \cos\left[\frac{(2y+1)j\pi}{2N}\right]$$

$$\boldsymbol{C}(\mathrm{u}) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{if u} = 0 \\ 1, & \text{if u} > 0 \end{cases}$$

Where p(x, y) is the x, yth element of the image represented by the matrix P. N is the size of the block. It is used to calculate the transformed image.

$$D(i,j) = \frac{1}{4} C(i)C(j) \sum_{x=0}^{7} \sum_{y=0}^{7} p(x,y) \cos\left[\frac{(2x+1)i\pi}{16}\right] \cos\left[\frac{(2y+1)j\pi}{16}\right]$$

For a standard 8*8 block image the value of the N is 8 and x, y varies from 0 to 7. DCT uses cosine functions; the resulting matrix depends on the horizontal, diagonal, and vertical frequencies. Therefore an image black with a lot of change in frequency has a very random looking resulting matrix, while an image matrix of just one color, has a resulting matrix of a large value for the first element and zeroes for the other elements.
Quantization is achieved by dividing each element in the transformed image matrix D by the corresponding element in the quantization matrix, and then round to the nearest integer value.

2.3 Feature Reduction with PCA

The features obtained with Gabor and DCT are at high dimensional space, so PCA is used to reduce the dimensions to lower dimensional space. Dimensionality reduction with PCA is performed with the following algorithm described in [4,8].

1. For a data matrix, $\mathbf{A_i}$ (i=1,…,6) of size M x N, where each column of **A** represented an image feature vector of length M (Γ_i) , N was the number of images representing each expression, and k was the facial expression index, zero-mean arrays $\widehat{A}_i$ were calculated by subtracting the mean values for each row:
$$m = \frac{1}{M} \sum_{i=1}^{M} \mathrm{A}i$$
2. Subtract the mean face from each face vector Γ_i to get a set of vectors φ_i. The purpose of subtracting the mean image from each image vector is to be left with only the distinguishing features from each face and "removing" in way information that is common. $\varphi_i = \Gamma\mathrm{i} - \mathrm{m}$
3. Find the Covariance matrix C; $C = A^T A$, where $A = [\varphi_1, \varphi_2 \ldots \varphi_M]$
4. Calculate the Eigenvectors μ_i of C
5. Select largest M Eigenvectors out of μ_i and ignore eigenvectors of small eigenvalues.

2.4 Wavelet Fusion

Since there are difficulties to obtain accurate and reliable results from a single feature, it is natural to adopt fusion techniques. Information fusion can be divided into three levels: data fusion, feature fusion and decision fusion [1, 10].

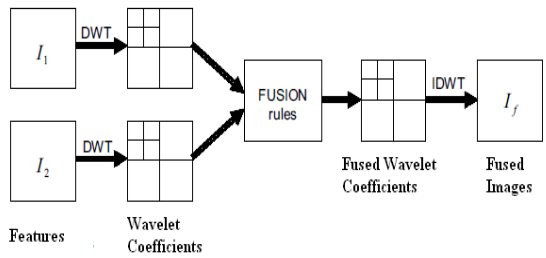

Fig 4: Feature Level Wavelet Fusion

In this study, wavelet fusion method is used to integrate all features into new combined features. This method is compared with different fusion methods such Maximum, Average, and PCA fusion [10]. The figure [4] shows the wavelet fusion method.

2.5 Classification with Neural Networks:

Radial Basis Function (RBF) networks have been used successfully as a classifier in many kinds of applications, including prediction, function approximation, and classification. It provides more robust and efficient capabilities than the conventional neural networks. A RBF neural network (RBFNN) is a universal approximator and its learning speed is very fast because of locally tuned neurons. RBF network is used to recognize facial emotions with extracted features. A RBFNN can be considered as a mapping: $R_r \rightarrow R_s$. Let $P \in R_r$ be the input vector and $Ci \in Rr$ be the prototype of the input vectors. The output of each RBF unit can be written as $R_i(P) = R_i(\|P - C_i\|)$ i=1…u

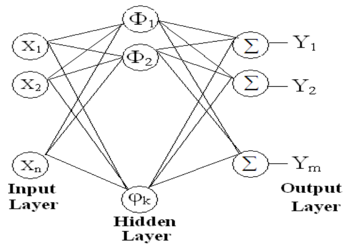

Fig 5: Basic Architecture of RBF Neural Network

The architecture of Probabilistic Neural Network (PNN) involves an input layer, a hidden layer and an output layer with feed forward architecture. The input layer of this network is a set of R units, which accept the elements of an R dimensional input feature vector. The input units are fully connected to the hidden layer with Q hidden units. Q is the number of input training pairs. Each target vector has K elements. One of these elements is 1 and rest are 0. Thus each input vector is associated with one of k classes. When an input vector is in the input layer, the hidden layer computes distances from the input vector to the training input vectors, and produces a vector whose elements indicate how close the input is to a training input. The output layer sums these contributions for each class of inputs to produce its net output as a vector of probabilities. Finally, a complete transfer function on the output of the output layer picks up the maximum of these probabilities and produces a 1 for that class and 0 for the other classes.

3 Experimental Results:

In this study, the static images of Cohn-Kanade[1] are used. Face area is detected and cropped from the raw faces and preprocessing is done to overcome illumination effects. Each image is cropped to size 200*200. In this database each image represents any of the most prominent emotion among six basic emotions. Three sample emotions of the Cohn-Kanade database is shown in the following figure. To validate proposed algorithms against variations of illumination and rotation, three different databases are generated with rotation angles of 2^o, 5^o and 10^o from standard benchmark database such as cohn-kanade facial expression database apart from original. Dataset1 is normalised original database. Dataset2 is generated with added rotational noise of -2^o . Dataset3 and 4 are generated with added rotational noise of -5^o and -10^o respectively. Samples images in cohn-kanade facial expression database for each expression is shown in figure 6.

Fig 6. Sample images for each expession

A total of 600 face images from 10 subjects were selected. The images were depicting six different facial expressions: surprise, disgust, happiness, fear, sadness and anger. In the training phase 180 images (3 for each expression) and in the testing phase 420 images were classified in scenario 1. The images used in the testing set were not included in the training set. The subjects represented in the training set were not included in the testing set of images, thus ensuring generalization of classification of facial emotions. The percentages of correct classifications are listed in the confusion tables (see Table 1,2 and 3). Table 1 shows the confusion between different facial expressions and the percentage of correct classifications of proposed algorithm under scenario1.

Expression	SURPRISE	FEAR	SAD	ANGER	DISGUST	HAPPY
SURPRISE	70	0	0	0	0	0
FEAR	8	62	0	0	0	0
SAD	1	0	67	2	0	0
ANGER	1	0	2	67	0	0
DISGUST	0	0	0	0	68	2
HAPPY	0	0	0	0	0	70
OVERALL	96.1905					

Table 1. Confusion matrix for the proposed algorithm with wavelet fusion(db2) in scenario 1.

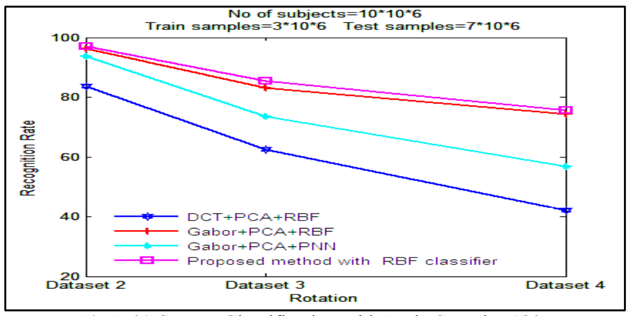

Fig 7: % Correct Classification with Train Samples 180

In scenario2, 240 images (4 for each expression) in the training phase are used to classify 360 images in the testing phase. Table 2 shows the confusion between different facial expressions and the percentage of correct classifications of proposed algorithm under scenario2.

Expression	SURPRISE	FEAR	SAD	ANGER	DISGUST	HAPPY
SURPRISE	60	0	0	0	0	0
FEAR	1	59	0	0	0	0
SAD	1	0	58	1	0	0
ANGER	0	0	1	58	1	0
DISGUST	0	0	0	0	60	0
HAPPY	0	0	0	0	0	60
OVERALL	98.6111					

Table 2: Confusion matrix for the proposed algorithm with wavelet fusion(db2) in scenario 2.

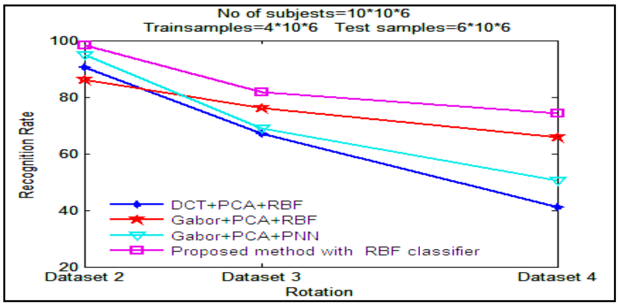

Fig 8: % of the correct classification with Train Samples 240

In scenario3, 300 images (5 for each expression) in the training phase are used to classify 300 images in the testing phase. Table 3 shows the confusion between different facial expressions and the percentage of correct classifications of proposed algorithm under scenario3.

Expression	SURPRISE	FEAR	SAD	ANGER	DISGUST	HAPPY
SURPRISE	50	0	0	0	0	0
FEAR	1	49	0	0	0	0
SAD	1	0	49	0	0	0
ANGER	0	0	1	48	1	0
DISGUST	0	0	0	0	50	0
HAPPY	0	0	0	0	0	50
OVERALL	98.66					

Table 3: Confusion matrix for the proposed algorithm with wavelet fusion(db2) in scenario 3.

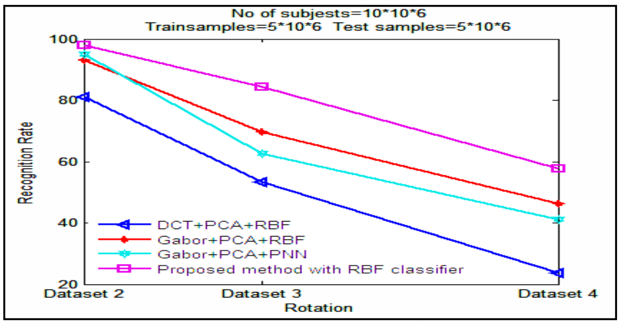

Fig 9: % of Correct Classification with proposed classifier for scenario 3.

The proposed method (Gabor+ DCT+ PCA+ RBF+ Wavelet Fusion) is compared with PNN classification having different fusion techniques. And the proposed method has highest recognition rate with low time complexity. In this scenario 180 samples (3 for each expression) has taken in the training phase and the remaining 420 samples are classified.

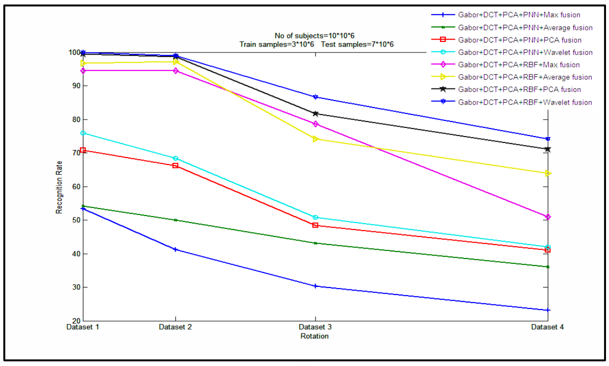

Fig 10: Recognition rate of proposed method with different fusion techniques.

Proposed method	Fusion Technique	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Gabor+ DCT +PCA+PNN	Maximum	53.33	41.22	30.27	23.11
	Average	54.11	50	43.11	36.11
	PCA	70.83	66.11	48.33	41.11
	Wavelet	75.88	68.33	50.83	41.94
Gabor+ DCT +PCA+RBF	Maximum	94.44	94.42	78.61	51
	Average	96.66	97.28	74.16	63.89
	PCA	99.44	98.61	81.66	71.11
	Wavelet	100	98.89	86.67	74.17

Table 4: Recognition Rates of proposed method with different classifiers on Dataset 1,2,3 and 4

4. Conclusion

A facial emotion recognition method with multi feature fusion using wavelet fusion was presented. Six different facial emotions were classified. The classification results were compared between different fusion techniques with the proposed technique. The correct classification rate was significantly high in proposed method comparative with other fusion methods. At the same time the recognition results were observed that it outperforms other existing single feature extraction methods. This indicates that different feature selection strategies would have to be implemented targeting specific facial expressions. The percentage of the correct classifications varied across different databases from 75% to 100%. Further studies are needed to develop better objectives and optimisation methods for feature selection and a novel neural network that would give higher recognition rates between similar facial expressions and under inclusion of noise in input images. The proposed method is also limited; the effectiveness of extraction expression feature is completely dependent on the effectiveness of pre-processing of the raw image. Further work involves taking many effective measures to improve the recognition accuracy.

References

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

[2] Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi, Jiro Gyoba, "Coding Facial Expressions with Gabor Wavelets" Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205, 1998.

[3] Donato, G., Bartlett, M.S., Hager, J.C., Ekman, P., Sejnowski, T.J. "Classifying Facial Actions". IEEE Trans.on Pattern Analysis Machine Intelligence. pp. 974-989, 1999.

[4] Matthem Turk and Alex Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience Volume 3, Number 1, 1991.

[5] Chien-Cheng Lee and Cheng-Yuan Shih, "Gabor Feature Selection and Improved Radial Basis Function Networks for Facial Expression Recognition", proc., by IEEE, 2009.

[6] Hadi Seyedarabi, Ali Aghagolzadeh, Sohrab Khanmohammadi "Recognition of Six Basic Facial Expressions by Feature-Points Tracking using RBF Neural Network and Fuzzy Inference System", IEEE International Conference on Multimedia and Expo (ICME), pp 1219-1222, 2004.

[7] Linlin Shen, Li Bai, "A review on Gabor wavelets for face recognition", proc., Pattern Anal Applic vol- 9, pp 273–292, 2006.

[8] Andrew J. Calder , A. Mike Burton , Paul Miller , Andrew W. Young , Shigeru kamatsu, A principal component analysis of facial expressions", Vision Research vol-41, pp 1179–1208, 2001.

[9] Prof. G. U. Kharat, Dr. S.V.Dudul" Neural Network Classifier for Human Emotion Recognition from Facial Expressions Using Discrete Cosine Transform", IEEE, pp 653-658, 2008.

[10] Gian Luca Marcialis and Fabio Roli, "Fusion of LDA and PCA for Face Verification", Proceedings of the International ECCV Workshop Copenhagen on Biometric Authentication, 2002.

K.V.Krishna Kishore received his M.Tech degree in Computer Science and Engineering from Andhra University, Visakhapatnam. He is now working as Professor and Head of Computer Science and Engineering Department in Vignan University. He got more than 19 years of teaching and research experience. His current research interests include Digital image processing, Biometrics, Face recognition, Facial emotional classifications, Neural Networks, Pattern Recognition and Medical Imaging. He has published more than 15 journal and conference papers in these areas. He has received best teacher award for four times in Vignan's Engineering College. He has Chaired National Youth festival for three times.

G.P.S. Varma received his B.Tech, M.Tech degrees in Computer Science and Engineering from Mangalore University and REC Warangal respectively. He got Ph.D Degree from Andhra University, Visakhapatnam. He is currently working as Professor and Head of Information Technology Department, S.R.K.R. Engineering College, Bhimavaram, He got more than 17 years of teaching and research experience. His current research interests include Software Engineering, Data ware Housing and Mining, Digital image processing, Neural Networks, Pattern Recognition and Bio Informatics. He has published more than 49 International and National Journal and conference papers in these areas. He has authored 8 books in the field of Computer Science and Engineering. He has Chaired National and International Conferences for Nine times.