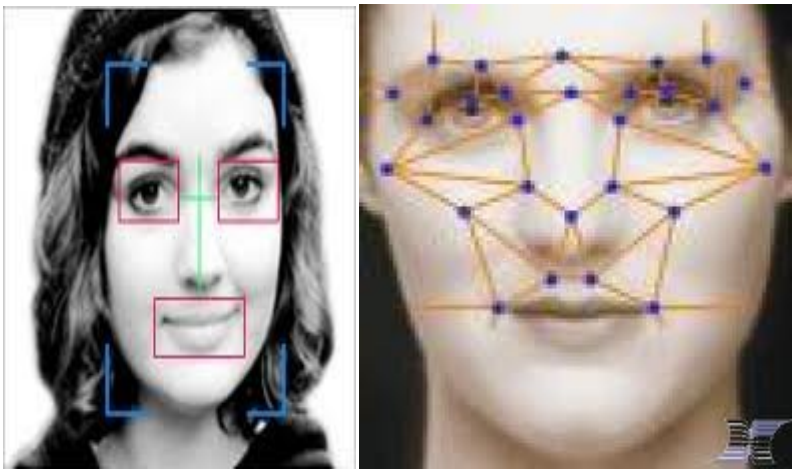
ADVANCED SURVEY ON FACE RECOGNITION TECHNIQUES IN IMAGE PROCESSING

NEERAJ SINGLA[1],SUGANDHA SHARMA[2]

Face recognition from image is a popular topic in biometrics research. the face recognition have played an important role in surveillance system as it doesn't need the object's cooperation. The actual advantages of face based identification over other biometrics are uniqueness and acceptance. As human face is a dynamic object having high degree of variability in its appearance, that makes face detection a difficult problem in computer vision. In this field, accuracy and speed of identification is a main issue. The goal of this paper is to evaluate various face detection and recognition methods, provide complete solution for image based face detection and recognition with higher accuracy, better response rate as an initial step for video surveillance. Solution is proposed based on performed tests on various face rich databases in terms of subjects, pose, emotions, race and light.

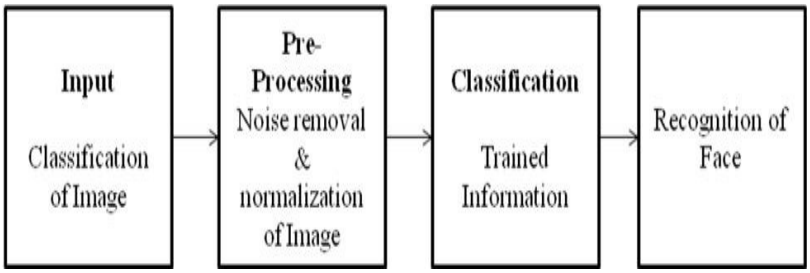
Keywords: *Face Detection, Face Recognition, Biometrics, Face Identification, emotions*



1. Introduction:

human emotions can be classified into six archetypal emotions: surprise, fear, disgust, anger, happiness, and sadness. Facial motion and the tone of the speech play a major role in expressing these emotions. The muscles of the face can be changed and the tone and the energy in the production of the speech can be intentionally modified to communicate different feelings. Human beings can recognize these signals even if they are subtly displayed, by simultaneously processing information acquired by ears and eyes. Based on psychological studies, which show that visual information modifies the perception of speech [17], it is possible to assume that human emotion perception follows a similar trend. some emotions are better identified with audio such as sadness and fear, and others with video, such as anger and happiness.

2. Face Detection:-



In the face detection, the input block stores the captured image wh ich finds the face area fro m the image. The face area provides to the pre-processing block which removes the unwanted noise and it also normalize the image. The output is provided to the trainer module, trains the image and decides whether the image belongs to the face class or not and finally it will provide the information about the recognition of face [1]. AdaBoost [6] classifier is used with Haar [7] and Local Binary Pattern (LBP) [8] features whereas Support Vector Machine (SVM) [12] classifier is used with Histogram of Oriented Gradients (HOG) [13] features for face detection evaluation. Haar-like [7] features are evaluated through the use of a new image representation that generates a large set of features and uses the boosting algorithm AdaBoost [6] to reduce degenerative tree of the boosted classifiers for robust and fast interferences only simple rectangular Haar-like [7] features are used that provides a number of benefits like sort of ad-hoc domain knowledge is implied as well as a speed increase over pixel based systems, suggestive to Haar [7] basis functions equivalent to intensity difference readings are quite easy to compute. Implementation of a system that used such features would

ISSN: 2278 – 1323

provide a feature set that was far too large, hence the feature set must be only restricted to a small number of critical features which is achieved by boosting algorithm, Adaboost [6]. The original LBP [8] operator labels the pixels of an image by thresholding the 3-by-3 neighborhood of each pixel with the center pixel value and considering the result as a binary number. Each face image can be considered as a composition of micro-patterns which can be effectively detected by the LBP [8] operator. To consider the shape information of faces, they divided face images into N small non-overlapping regions T_0, T_1, ..., T_N. The LBP [8] histograms extracted from each sub-region are then concatenated into a single, spatially enhanced feature histogram defined as:

$$H_{i,j} = \Sigma_{x,y} I(f_l(x,y)=i) I((x,y) \; T_j)$$

where i = 0, ... , L-1; j = 0, ..., N-1. The extracted feature histogram describes the local texture and global shape of face images.

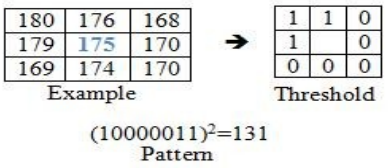


Fig. 2 LBP calculation.

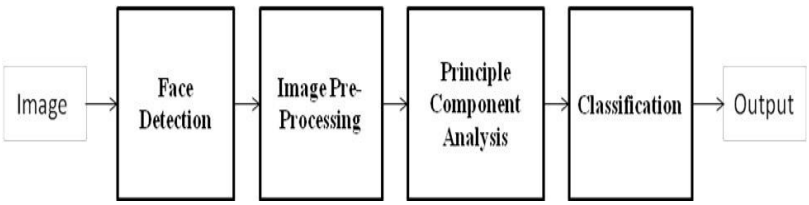
SVM [12] classifier is been used with HOG [13] features for face detection. HOG [13] greatly outperforms wavelets and degree of smoothing before calculating gradients damages, results emphasizes much of the available information is from sudden edges at fine scales that blurring this for reducing the sensitivity to spatial position is a mistake. Gradients should be calculated at the finest available scale in the current pyramid layer and strong local contrast normalization is essential for good results. Whereas SVM [12] are formulated to solve a classical two class problem which returns a binary value, the class of the object. To train our SVM [12] algorithm, we formulate the problem in a difference space that explicitly captures the dissimilarity between two facial images. The results summery of above methods are stated below.

Moreover system is been tested on datasets [1,2,3,4,5] and based on above demonstrated system, results are demonstrated below:

To reduce pose variation and illumination in extracted faces two extra actions performed in pre-processing stage to improve recognition results: 1) Eyes detection is been

ISSN: 2278 – 1323

2. RECOGNITION SYSTEMS: In this article the basic system proposed four stages: face detection, pre-processing, principle componenet anaysis (PCA) and clas sification.



2.1 Emotion recognition by facial expressions

Facial expressions give important clues about emotions. Therefore, several approaches have been proposed to classify human affective states. The features used are typically based on local spatial position or displacement of specific points and regions of the face, unlike the approaches based on audio, which use global statistics of the acoustic features. Mase proposed an emotion recognition system that uses the major directions of specific facial muscles . With 11 windows manually located in the face, the muscle movements were extracted by the use of optical flow. For classification, K-nearest neighbor rule was used, with an accuracy of 80% with four emotions: happiness, anger, disgust and surprise. Yacoob et al. proposed a similar method Instead of using facial muscle actions, they built a dictionary to convert motions associated with edge of the mouth, eyes and eyebrows, into a linguistic, per-frame, mid-level representation. Black et al. used parametric models to extract the shape and movements of the mouse, eye and eyebrows [1]. They also built a mid- and high-level representation of facial actions by using a similar approach employed in [22], with 89% of accuracy. Tian et al. attempted to recognize Actions Units (AU), developed by Ekman and Friesen in 1978 [10], using permanent and transient facial features such as lip, nasolabial furrow and wrinkles [21]. Geometrical models were used to locate the shapes and appearances of these features. They achieved a 96% of accuracy. Essa et al. developed a system that quantified facial movements based on parametric models of independent facial muscle groups [11]. They modeled the face by the use of an optical flow method coupled with geometric, physical and motion-based dynamic models. They generated spatial-temporal templates that were used for emotion recognition. Without considering sadness that was not included in their work, a recognition accuracy rate of 98% was achieved.

2.2 Emotion recognition by bimodal data

Relatively few efforts have focused on implementing emotion recognition systems using both facial expressions and acoustic information. De Silva et al. proposed a rule-based audio-visual emotion recognition system, in which the outputs of the uni-modal classifiers are fused at the decision-level [8]. From audio, they used prosodic features, and from video, they used the maximum distances and velocities between six specific facial points. A similar approach was also presented by Chen et al. [4], in which the dominant modality, according to the subjective experiments conducted in [7], was used to resolve discrepancies between the outputs of mono-modal systems. In both studies, they concluded that the performance of the system increased when both modalities were used together.Yoshitomi et al. proposed a multimodal system that not only considers speech and visual information, but also the thermal distribution acquired by infrared camera [24]. They argue that infrared images are not sensitive to lighting conditions, which is one of the main problems when the facial expressions are acquired with conventional cameras. They used a database recorded from a female speaker that read a single word acted in five emotional states. They integrated these three modalities at decision-level using empirically determined weights. The performance of the system was better when three modalities were used together. a bimodal emotion recognition system was proposed to recognize six emotions, in which the audio-visual data was fused at feature-level. They used prosodic features from audio, and the position and movement of facial organs from video. The best features from both unimodal systems were used as input in the bimodal classifier.

3. METHODOLOGY

Four emotions -- sadness, happiness, anger and neutral state --are recognized by the use of three different systems based on audio, facial expression and bimodal information, respectively. The main purpose is to quantify the performance of unimodal systems, recognize the strengths and weaknesses of these approaches and compare different approaches to fuse these dissimilar modalities to increase the overall recognition rate of the system.

The database used in the experiments was recorded from an actress who read 258 sentences expressing the emotions. A VICON motion capture system with three cameras (left of Figure 1) was used to capture the expressive facial motion data with 120Hz sampling frequency. With 102 markers on her face (right of Figure 1), an actress was asked to speak a custom phoneme-balanced corpus four times, with different emotions. The recording was made in a quiet room using a close talking SHURE microphone at the sampling rate of 48 kHz. The markers' motion and aligned audio were captured by the system simultaneously. Notice that the facial features are extracted with high precision, so this multimodal database is suitable to extract important clues about both facial expressions and speech.

Data recording system

In order to compare the unimodal systems with the multimodal system, three different approaches were

ISSN: 2278 – 1323

implemented all using support vector machine classifier (SVC) with 2 nd order polynomial kernel functions [3]. SVC was used for emotion recognition in our previous study, showing better performance than other statistical classifiers [13][14]. Notice that the difference between the three approaches is in the features used as inputs, so it is possible to conclude the strengths and limitations of acoustic and facial expressions features to recognize human emotions. In all the three systems, the database was trained and tested using the leave-one-out cross validation method.

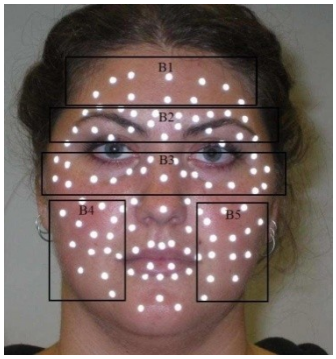
3.1 System based on speech

The most widely used speech cues for audio emotion recognition are global-level prosodic features such as the statistics of the pitch and the intensity. Therefore, the means, the standard deviations, the ranges, the maximum values, the minimum values and the medians of the pitch and the energy were computed using Praat speech processing software [2]. In addition, the voiced/speech and unvoiced/speech ratio were also estimated. By the use of sequential backward features selection technique, a 11-dimensional feature vector for each utterance was used as input in the audio emotion recognition system.

3.2 System based on facial expressions

In the system based on visual information, which is described in figure 4, the spatial data collected from markers in each frame of the video is reduced into a 4-dimensional feature vector per sentence, which is then used as input to the classifier. The facial expression system, which is shown in figure 4, is described below.

After the motion data are captured, the data are normalized: (1) all markers are translated in order to make a nose marker be the local coordinate center of each frame, (2) one frame with neutral and close-mouth head pose is picked as the reference frame, (3) three approximately rigid markers (manually chosen and illustrated as blue points in Figure 1) define a local coordinate origin for each frame, and (4) each frame is rotated to align it with the reference frame. Each data frame is divided into five blocks: forehead, eyebrow, low eye, right cheek and left cheek area (see Figure 2). For each block, the 3D coordinate of markers in this block is concatenated together to form a data vector. Then, Principal Component Analysis (PCA) method is used to reduce the number of features per frame into a 10-dimensional vector for each area, covering more than 99% of the variation. Notice that the markers near the lips are not considered, because the articulation of the speech might be recognized as a smile, confusing the emotion recognition system [19].



five areas of the face considered in this study

In order to visualize how well these feature vectors represent the emotion classes, the first two components of the low eye area vector were plotted in figure 3. As can be seen, different emotions appear in separate clusters, so important clues can be extracted from the spatial position of these 10-dimensional features space.

Notice that for each frame, a 10- dimensional feature vector is obtained in each block. This local information might be used to train dynamic models such as HMM. However, in this paper we decided to use global features at utterance level for both unimodal systems, so these feature vectors were preprocessed to obtain a low dimensional feature vector per utterance. In each of the 5 blocks, the 10-dimensional features at frame level were classified using a K-nearest neighbor classifier (k=3), exploiting the fact that different emotions appear in separate clusters (Figure 3). Then, the number of frames that were classified for each emotion was counted, obtaining a 4-dimentional vector at utterance level, for each block. These feature vectors at utterance level take advantage not only of the spatial position of facial points, but also of global patterns shown when emotions are displayed. For example, when happiness is displayed in more than 90 percent of the frames, they are classified as happy, but when sadness is displayed even more than 50 percent of the frames, they are classified as sad. The SVC classifiers use this kind of information, improving significantly the performance of the system. Also, with this approach the facial expression features and the global acoustic features do not need to be synchronized, so they can be easily combined in a feature-level fusion.

3.3 Bimodal system

To fuse the facial expression and acoustic information, two different approaches were implemented: feature-level fusion, in which a single classifier with features of both modalities are used (left of Figure 5); and, decision level fusion, in which a separate classifier is used for each modality, and the outputs are combined using some criteria (right of Figure 5). In the first approach, a sequential backward feature selection technique was used to find the features from both modalities that maximize the performance of the classifier. The number of features selected was 10. In the second approach, several criteria were used to combine the posterior probabilities of the mono-modal systems at the decision-level: maximum, in which the emotion with greatest posterior probability in both modalities is selected; average

4. RESULTS

4.1 Acoustic emotion classifier

Table 1 shows the confusion matrix of the emotion recognition system based on acoustic information, which gives details of the strengths and weaknesses of this system. The overall performance of this classifier was 70.9 percent. The diagonal components of table 1 reveal that all the emotions can be recognized with more than 64 percent of accuracy, by using only the features of the speech. However, table 1 shows that some pairs of emotions are usually confused more.

ISSN: 2278 – 1323

Sadness is misclassified as neutral state (22%) and vise versa (14%). The same trend appears between happiness and anger, which are mutually confused (19% and 21%, respectively) . These results agree with the human evaluations done by De Silva et al. [7], and can be explained by similarity patterns observed in acoustic parameters of these emotions [23]. For example, speech associated with anger and happiness is characterized by longer utterance duration, shorter inter-word silence, higher pitch and energy values with wider ranges. On the other hand, in neutral and sad sentences, the energy and the pitch are usually maintained at the same level. Therefore, these emotions are difficult to be classified.

Table 1: Confusion matrix of the emotion recognition system based on audio

	Anger	Sadness	Happiness	Neutral
Anger	0.68	0.05	0.21	0.05
Sadness	0.07	0.64	0.06	0.22
Happiness	0.19	0.04	0.70	0.08
Neutral	0.04	0.14	0.01	0.81

4.2 System based on facial expressions

Table 3 shows the performance of the emotion recognition systems based on facial expressions, for each of the five facial blocks and the combined facial expression classifier. This table reveals that the cheek areas give valuable information for emotion classification. It also shows that the eyebrows, which have been widely used in facial expression recognition, give the poorest performance. The fact that happiness is classified without any mistake can be explained by the figure 3, which shows that happiness is separately clustered in the 10-dimentional PCA spaces, so it is easily to recognize. Table 2 also reveals that the combined facial expression classifier has an accuracy of 85%, which is higher than most of the 5 facial blocks classifiers. Notice that this database was recorded from a *single* actress, so clearly more experiments should be conducted to evaluate these results with other subjects.

Table 2: Performance of the facial expression classifiers

Area	Overall	Anger	Sadness	Hapiness	Neutral
Forehead	0.73	0.82	0.66	1.00	0.46
Eyebrow	0.68	0.55	0.67	1.00	0.49
Low eye	0.81	0.82	0.78	1.00	0.65
Right cheek	0.85	0.87	0.76	1.00	0.79
Left cheek	0.80	0.84	0.67	1.00	0.67
Combined classifier	0.85	0.79	0.81	1.00	0.81

The combined facial expression classifier can be seen as a feature-level integration approach in which the features of the 5 blocks are fused before classification. These classifiers can be also integrated at decision-level. Table 3 shows the performance of the system when the facial block classifiers are fused by the use of different criteria. In general, the results are very similar. All these decision-level rules give slightly worse performance than the combined facial expression classifier.

Table 3: Decision-level integration of the 5 facial blocks emotion classifiers

	Overall	Anger	Sadness	Hapiness	Neutral
Majority voting	0.82	0.92	0.72	1.00	0.65
Maximum	0.84	0.87	0.73	1.00	0.75
Averaging combining	0.83	0.89	0.72	1.00	0.70
Product combining	0.84	0.87	0.72	1.00	0.77

Table 4 shows the confusion matrix of the combined facial expression classifier to analyze in detail the limitation of this emotion recognition system. The overall performance of this classifier was 85.1 percent. This table reveals that happiness is recognized with very high accuracy. The other three emotions are classified with 80 percent of accuracy, approximately. Table 4 also shows that in the facial expressions domain, anger is confused with sadness (18%) and neutral state is confused with happiness (15%) . Notice that in the acoustic domain, sadness/anger and neutral /happiness can be separated with high accuracy, so it is expected that the bimodal classifier will give good performance for anger and neutral state. This table also shows that sadness is confused with neutral state (13%). Unfortunately, these two emotions are also confused in the acoustic domain (22%), so it is expected that the recognition rate of sadness in the bimodal classifiers will be poor. Other discriminating information such as contextual cues are needed.

Table 4: Confusion matrix of the combined facial expression classifier

	Anger	Sadness	Happiness	Neutral
Anger	0.79	0.18	0.00	0.03
Sadness	0.06	0.81	0.00	0.13
Happiness	0.00	0.00	1.00	0.00
Neutral	0.00	0.04	0.15	0.81

4.3 Bimodal system

Table 5 displays the confusion matrix of the bimodal system when the facial expressions and acoustic information were fused at feature-level. The overall performance of this classifier was 89.1 percent. As can be observed, anger, happiness and neutral state are recognized with more than 90 percent of accuracy. As it was expected, the recognition rate of anger and neutral state was higher than unimodal systems. Sadness is the emotion with lower performance, which agrees with our previous analysis. This emotion is confused with neutral state (18%), because none of the modalities we considered can accurately separate these classes. Notice that the performance of happiness significantly decreased to 91 percent.

ISSN: 2278 – 1323

Table 5: Confusion matrix of the feature-level integration bimodal classifier

	Anger	Sadness	Happiness	Neutral
Anger	0.95	0.00	0.03	0.03
Sadness	0.00	0.79	0.03	0.18
Happiness	0.02	0.00	0.91	0.08
Neutral	0.01	0.05	0.02	0.92

Table 6 shows the performance of the bimodal system when the acoustic emotion classifier (Table 1) and the combined facial expressions classifier (Table 4) were integrated at decision-level, using different fusing criteria. In the weight-combining rule, the modalities are weighted according to rules extracted from the confusion matrices of each classifier. This table reveals that the maximum combining rule gives similar results compared to the facial expression classifier. This result suggests that the posterior probabilities of the acoustic classifier are smaller than the posterior probabilities of the facial expression classifier. Therefore, this rule is not suitable for fusing these modalities, because one modality will be effectively ignored. Table 6 also shows that the product-combining rule gives the best performance.

Table 6: Decision-level integration bimodal classifier with different fusing criteria

	Overall	Anger	Sadness	Hapiness	Neutral
Maximum combining	0.84	0.82	0.81	0.92	0.81
Averaging combining	0.88	0.84	0.84	1.00	0.84
Product combining	0.89	0.84	0.90	0.98	0.84
Weight combining	0.86	0.89	0.75	1.00	0.81

Table 7 shows the confusion matrix of the decision-level bimodal classifier when the product-combining criterion was used. The overall performance of this classifier was 89.0 percent, which is very close to the overall performance achieved by the feature-level bimodal classifier (Table 5). However, the confusion matrices of both classifiers show important differences. Table 7 shows that in this classifier, the recognition rate of anger (84%) and neutral states (84%) are slightly better than in the facial expression classifier (79% and 81%, Table 4), and significantly worse than in the feature-level bimodal classifier (95%, 92%, Table 5). However, happiness (98%) and sadness (90%) are recognized with high accuracy compared to the feature-level bimodal classifier (91% and 79%, Table 5). These results suggest that in the decision-level fusion approach, the recognition rate of each emotion is increased, improving the performance of the bimodal system.

Table 7: Confusion matrix of the decision-level bimodal classifier with product-combining rule

	Anger	Sadness	Happiness	Neutral
Anger	0.84	0.08	0.00	0.08
Sadness	0.00	0.90	0.00	0.10
Happiness	0.00	0.00	0.98	0.02
Neutral	0.00	0.02	0.14	0.84

6. CONCLUSION

This research analyzed the strengths and weaknesses of facial expression classifiers and acoustic emotion classifiers. In these unimodal systems, some pairs of emotions are usually misclassified. However, the results presented in this paper show that most of these confusions could be resolved by the use of another modality. Therefore, the performance of the bimodal emotion classifier was higher than each of the unimodal systems Two fusion approaches were compared: feature-level and decision-level fusion. The overall performance of both approaches was similar. However, the recognition rate for specific emotions presented significant discrepancies. In the feature-level bimodal classifier, anger and neutral state were accurately recognized compared to the facial expression classifier, which was the best unimodal system. In the decision-level bimodal classifier, happiness and sadness were classified with high accuracy.

REFERENCES

[1] Black, M. J. and Yacoob, Y. Tracking and recognizing rigid and non-rigid facial motions using local parametric model of image motion. In Proceedings of the International Conference on Computer Vision, pages 374–381. IEEE Computer Society, Cambridge, MA, 2012.

[2] Boersma, P., Weenink, D., Praat Speech Processing Software, Institute of Phonetics Sciences of the University of Amsterdam,2010.

[3] Burges, C. A tutorial on support vector machines for pattern recognition. Dat Mining and Know. Disc., vol. 2(2), pp. 1– 47, 2006.

[4] Chen, L.S., Huang, T. S., Miyasato T., and Nakatsu R. Multimodal human emotion / expression recognition, in Proc. of Int. Conf. on Automatic Face and Gesture Recognition, (Nara, Japan), IEEE Computer Soc., April 2009.

[5] Chen, L.S., Huang, T.S. Emotional expressions in audiovisual human computer interaction. Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on, Volume: 1, 30 ,Pages: 423 - 426 ,2011.

[6] De Silva, L.C., Ng, P. C. Bimodal emotion recognition. Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, 28-30 ,Pages: 332 – 335,2011.

[7] Dellaert, F., Polzin, T., Waibel, A. Recognizing emotion in speech. Spoken Language, 1996. ICSLP 96. Proceedings. Fourth International Conference on, Volume: 3, Pages: 1970 - 1973 vol.3,2011.

[8] Ekman, P., Friesen, W. V. Facial Action Coding System: A Technique for Measurement of Facial Movement. Consulting Psychologists Press Palo Alto, California, 2011

ISSN: 2278 – 1323

BIOGRAPHY

ER. NEERAJ SINGLA, pursuing M.Tech - CSE from CGC, Gharuan and done B.Tech. degree from Punjab technical university.he is the author o 5 international journal.

ER. SUGANDHA SHARMA is working as Assistant professor in CGC Gharuan in CSE department .. She completed her M.E in Computer Science and Engineering from University Institute of Engineering and Technology, Panjab University in 2010. She is Author of EIGHT International Journals. Fields of Specialization is Image ProcessinG.

ISSN: 2278 – 1323