A Survey on various objective Image Quality Assessment Techniques

Raman Gupta, Dipti Bansal, Charanjit Singh

***Abstract*— Image quality assessment enables to approximate the quality of an image and is used in number of image processing applications. Quality of an image can be measured in two ways: subjective IQA and objective IQA method. Objective method is considered to be better than subjective method because most of the time the reference image is not accessible for the comparison. Also, objective method is cheaper than the subjective method. These methods are used to calculate the visual quality by linking a distorted image against original image. In this paper we are comparing the various approaches of image quality assessment.**

***Index Terms*— Image Quality Assessment, Mean Opinion Score, Human Visual System.**

I. Introduction

Image quality assessment is one among the emergent field of digital image processing. Many researchers are working on the parameters that affect the image quality. They would like to know how to attain an imaging system that achieves a particular level of image quality at the lowest possible cost. Improvement of image processing has been powered by advancement in technologies such as, expansion in digital images, computer processors, mass storage devices, etc. Number of fields which commonly used analog imaging are now switching to digital systems, for their affordability and flexibility. Few examples are film and video production, medicine, remote sensing, photography and security monitoring, etc. [1]. Digital image processing is mainly concerned with extracting useful information from images. Use of image quality metric play an important role for the following application [2]:

- To monitor image quality for quality control system. For example, quality of digital video transmitted on a network is examined by network video server.
- Benchmarking an image processing system and algorithms. For example, quality metric is used to select one from multiple image processing systems which provide the best quality images.
- Optimizing the algorithms and the parameter setting of an image processing. For example, a quality metric is used for optimal design of the pre-filtering algorithms at the encoder and post-filtering algorithms at the decoder.

Image quality assessment is done in two ways: Subjective method and Objective method. The ultimate goal of quantifying the visual quality is to get the opinion of human observers, known as subjective quality evaluation in which mean opinion score (MOS) is evaluated. This method has been widely used for many years. But in practical usage, the MOS method is inconvenient, expensive and time consuming. Thus objective image quality metrics are preferred and the goal of which is to supply quality metrics that can predict perceived image quality automatically. The most widely used objective image quality metrics are peak signal-to-noise ratio (PSNR) and mean square error (MSE). Although they are computationally simple, they does not correlate well with the perceived quality measurement, thus they are widely criticized [2]. A great deal of effort has been made to design new objective quality assessment methods that are consistent with perceptual quality measures.
All these methods want to have high correlation with human perception or judgments. In this paper we are reviewing various methods used to assess image quality.

II. Subjective Image Quality Assessment

In subjective quality assessment, image are provided to a number of observers and are asked to compare original images with distorted images in order to evaluate the quality of the distorted images. Based on their evaluation, mean opinion score (MOS) is calculated which is taken as the image quality index [3]. No mathematical equation is used in subjective method. This method is considered costly, inconvenient and time consuming.

Three factors: luminance, viewing distance from observer to display and display properties are taken into account while conducting the subjective quality test. For subjective assessment of image quality, at least 15 observers should be used and at least four different types of scenes should be chosen.

A. Double stimulus impairment scale (DSIS)

The DSIS is used to evaluate the degradation level of the distorted image with respect to the original image. In this method, first each observer views an unimpaired original image and then its impaired version. Observers are then asked to vote on the second, keeping first in mind using a scale containing 5 scores [4]: Imperceptible (5), Perceptible but not annoying (4), Slightly annoying (3), Annoying (2), Very annoying (1).

B. Double-Stimulus Continuous Quality-Scale (DSCQS)

The DSCQS method is primarily used when it is not

Manuscript received July 19, 2014

Raman Gupta, Department of Electronics and Communication Engineering, Punjabi University, Patiala, Punjab, India.
Er. Dipti Bansal, Department of Electronics and Communication Engineering, Punjabi University, Patiala, Punjab, India.
Dr. Charanjit Singh, Department of Electronics and Communication Engineering, Punjabi University, Patiala, Punjab, India.

possible to provide test stimulus test conditions that exhibit full range of quality. In this method, the observer is asked to view a pair of visual sequence. A pair consist of one image via the process under examination and other directly from the source. Observers then vote on the both image. If the observer is alone, is allowed to trigger between the original image and test sequences until opinion on each image is established [4]. Otherwise, if multiple are evaluating simultaneously, they are shown reference and test sequences twice to make their opinion of each.

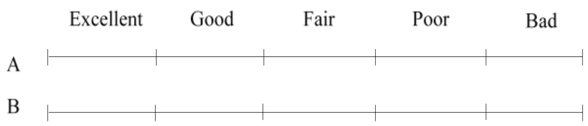

Fig. 1: Quality rating form

In order to rate the quality of both images, the dual vertical scale is used. The scale is distributed into five equal lengths which relates to quality scales as shown in Fig. 1.

C. Single Stimulus Continuous Quality Scale (SSCQS)

The SSCQE approach is useful to evaluate digitally coded video which is scene-dependent and have time-varying impairments. In this technique image sequences without a reference are presented to the observer only once Observers continuously weigh the image sequence along the time on a linear scale by an electronic recording handset associated to a computer [4] and provide a result as ‘good’ or ‘bad’.

D. Simultaneous Double Stimulus for Continuous Evaluation (SDSCE)

The SDSCE scheme is appropriate where fidelity of pictorial information affected by time-varying degradation has to be assessed. In this technique image sequences are offered in pairs such that original and impaired sequences are presented side by side at same time. Then, the observers are enquired to check the alterations amid the two sequences and to evaluate the fidelity of the image information along the time on a linear scale by an electronic recording handset attached to a computer. The observers are conscious of the original and distorted sequences throughout calculation session. After the calculation session, data is collected from the tests and processed to achieve a level of impairment.

III. Objective Quality Assessment

In objective quality assessment, automatic algorithms or mathematical equations are used for quality assessment that could analyze images and report their quality without human involvement. This method reduce the cost and make quality assessment faster. Based on the availability of an original image, objective image quality metrics are classified as [3]:-

- Full-reference: when complete reference image is assumed to be known.
- No-reference: when reference image is not available. This is also known as “blind quality assessment”.
- Reduced-reference: when reference image is known partially in the form of a set of extracted features as side information that helps in evaluation.

Full reference image quality measures is again classified into six classes of objective quality measures [3]:

- Pixel difference-based measures: It includes mean square error (MSE), signal-to-noise ratio (SNR) and peak signal-to-noise ratio (PSNR). These metrics are easy to evaluate.
- Correlation-based measures: it measures the difference between two digital images. Correlation of pixel is used to measure the image quality in image quality assessment.
- Edge-based measure: in this class, relative displacement of edge positions between reference image and distorted image or there consistency are used to evaluate the image quality.
- Spectral distance-based measures: in this objective measure, Discrete Fourier Transform is applied on the reference and the distorted image and their difference of the Fourier magnitude or phase spectral is treated as an image quality measure.
- Context-based measures: in this class, neighboring pixels are compared against each other by finding the multidimensional contest probability that is used to measure image quality.

Human Visual System-based measures: this image quality measure is based on the perception of the human eyes which usually use contrast, color and frequency changes in their measures.

IV. Full Reference Image Quality Assessment

A. Peak Signal to Noise Ratio (PSNR)

Peak signal to noise ratio, often abbreviated as PSNR, is an engineering term that gives the ratio between the maximum power present in the image and power of the corrupting noise present in that same image. This ratio is used as a quality measurement between the original and a compressed image.

PSNR can be easily defined by using Mean Squared Error (MSE) which is given as,

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I(i,j)-K(i,j)\right]^2 \tag{1}$$

where, m is the picture width, n is the picture height, $I(i, j)$ is the original frame at pixel position (i, j) and $K(i, j)$ is the distorted frame at pixel position (i, j).

Using MSE, PSNR can be defined as [5]:

$$PSNR = 10.\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{2}$$

where, MAX_I is the maximum number of pixels in the image. Although it is computationally simple and widely used in image and video quality evaluation, it does not correlate with the subjective evaluation.

B. Structural Similarity (SSIM)

SSIM refers structural similarity that is used for measuring similarity between two images. The SSIM metric is full reference engineering approach in which initial uncompressed or distortion-free image is used as reference. In SSIM, an assumption is taken that HVS (human visual system) is highly sensitive to structural distortions [6].

For image quality assessment using SSIM, a system is made which separates the task of similarity measurement into three comparisons: luminance comparison, contrast comparison and structure comparison. After combining the three comparisons, the overall similarity measure is defined as:

$$S(x,y) = f\left(l(x,y), c(x,y), s(x,y)\right) \quad (3)$$

where, $l(x,y)$, $c(x,y)$ and $s(x,y)$ are comparison functions and $f(.)$ is the combination function.

C. Multi-scale structural similarity (MSSIM)

MSSIM (Multi-scale structural similarity) is the extension of single-scale SSIM. It provides more flexibility to incorporate the variations of viewing conditions than previous single-scale method [7]. Viewing conditions are taken into account before moving to a multi-scale approach. These viewing conditions are display resolution and viewing distance.

The original and distorted image signal passed through a low-pass filter which down-samples the filtered image by 2 iteratively. The scaling of the image is done from scale 1 to scale *M* which is obtained after *M*-1 iterations. At the j^{th} scale the structure comparison and contrast comparison are evaluated and represented as $s_j(x,y)$ and $c_j(x,y)$ respectively. The luminance comparison $I_M(x,y)$ is evaluated only at scale *M*. By combining the measurements at different scales, an overall MSSIM index is evaluated as given below:-

$$MSSIM(x,y) = \left[l_M(x,y)\right]^{\alpha_M} \cdot \prod_{j=1}^{M} \left[c_j(x,y)\right]^{\beta_j} \left[s_j(x,y)\right]^{\gamma_j} \quad (3)$$

where, α_M, β_j and γ_j denotes the parameters that are used to adjust relative importance of different components [8].

D. Visual Information Fidelity (VIF)

VIF (visual information fidelity) is based on the relationship between image information and visual quality. It is full-reference vision modeling approach in which the two quantities, which are: the information in the original image and how much of this original information can be extracted from the test image, are combined. In VIF measure as purposed in [9], original image is taken as the output of a stochastic "natural" source. This signal is then passed through the human visual system (HVS) and then enters the brain for processing. The original signal has passed through distortion channel before entering the HVS. The VIF is derived by quantifying two mutual information quantities: first is mutual information between the input and the output of the HVS channel and other one is the mutual information between the input of the distortion channel and the output of the HVS channel for the test image [9]. Fig. 2 shows the relation pictorially.

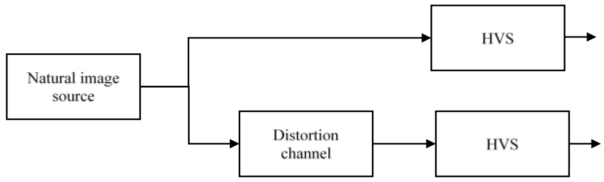

Fig. 2: Source, distortion and HVS model relationship

E. Feature Similarity (FSIM)

FSIM (Feature Similarity) is a full reference image quality assessment method which is based on fact that human visual system understands an image according to its low-level features [10]. To find the FSIM index two features, Phase congruency (PC) and Gradient magnitude (GM) are to be evaluated. PC is used as a primary feature in FSIM and it is dimensionless measure of the significance of a local structure. GM is considered to be a second feature. PC and GM are complementary in characterizing the image local quality.

The evaluation of FSIM index is done in two steps. First, local similarity map is calculated and then the similarity mat is pooled into a single similarity score. The FSIM measurement is separated between $f_1(x)$ and $f_2(x)$ into two components, each for *PC* or *GM*. The similarity measure in terms of $PC_1(x)$ and $PC_2(x)$ is defined as:

$$S_{PC}(x) = \frac{2PC_1(x) \cdot PC_2(x) + T_1}{PC_1^2(x) + PC_2^2(x) + T_1} \quad (4)$$

where, T_1 is a positive constant used to increase the stability of S_{PC}. Similarly, the similarity measure in terms of *GM* values $G_1(x)$ and $G_2(x)$ is given as:

$$S_G(x) = \frac{2G_1(x) \cdot G_2(x) + T_2}{G_1^2(x) + G_2^2(x) + T_2} \quad (5)$$

Then, the FSIM index between f_1 and f_2 can be defined as:-

$$FSIM = \frac{\sum_{x \in \beta} S_L(x) \cdot PC_m(x)}{\sum_{x \in \beta} PC_m(x)} \quad (6)$$

where, β denotes whole image spatial domain.

F. Universal Image Quality (UQI)

UQI (Universal image quality) Index is an objective image quality assessment index that is easy to compute. UQI can be calculated by modeling any image distortion as a combination of three factors [11]:-

- Loss of correlation
- Luminance distortion
- Contrast distortion

This index is independent of viewing conditions and individual observers. If $x = \{x_i \mid i = 1, 2, \ldots., N\}$ be the reference image signal and $y = \{y_i \mid i = 1, 2, \ldots.., N\}$ be the test image signal, then Universal Image Quality Index can be defined as:-

$$Q = \frac{4\sigma_{xy}\overline{x}\overline{y}}{\left(\sigma_x^2+\sigma_y^2\right)\left[\left(\overline{x}\right)^2+\left(\overline{y}\right)^2\right]} \quad (7)$$

where,

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N}x_i, \quad \overline{y} = \frac{1}{N}\sum_{i=1}^{N}y_i, \quad \sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i-\overline{x}\right)^2$$

$$\sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i-\overline{y}\right)^2, \sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)$$

The dynamic range of Q is [1, -1].

G. *Peak signal to noise ratio-human visual system (PSNR-HVS)*

PSNR-HVS (Peak signal to noise ratio- human visual system) is a full reference metrics for computing the PSNR while taking into account the HVS as HVS is more sensitive to low frequency distortions than high frequency distortions [12]. The flow chart for the calculation of PSNR-HVS is shown below:

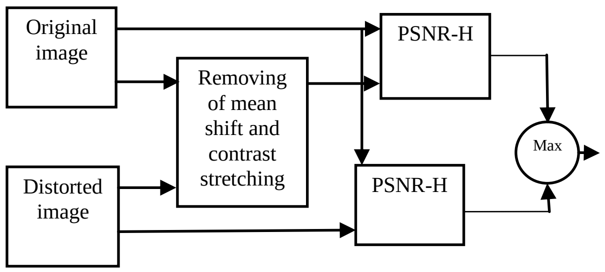

Fig. 3: PSNR-HVS System

If window size used is 64×64 pixels, then PSNR-HVS is given as:

$$PSNRHVS = 10\log\left\{\frac{255^2}{MSE_H}\right\} \quad (8)$$

In above equation, MSE_H is calculated taking HVS into account and given as:-

$$MSE_H = K\sum_{i=1}^{I-7}\sum_{j=1}^{J-7}\sum_{m=1}^{8}\sum_{n=1}^{8}\left(\left(X\left[m,n\right]_{ij} - X\left[m,n\right]_{ij}^{e}\right)T_c\left[m,n\right]\right)^2 \quad (9)$$

where, (I, J) is image size, $K = 1/\left[\left(I-7\right)\left(J-7\right)64\right]$, X_{ij} are DCT coefficients of 8×8 image block, X_{ij}^{e} are DCT coefficients of the corresponding block in the original image, and T_c represents the metric of correcting factors.

V. Image Database

To assess the performance of objective quality metrics, it is essential to obtain database of test images from which subjective quality score (Mean Opinion Score) has been experimentally collected. TID2013 (Tampere Image Database 2013) [13], which is publicly-available database is used for this purpose. The TID2013 comprises 25 reference images, 24 types of distortions for each reference image, and 5 dissimilar levels for each type of distortion. The entire database enclose 3000 distorted images. Reference images are attained by cropping from Kodak Lossless True Color image suite and kept them in database in Bitmap setup without any compression. File name of each image specify a number of the reference image, a number of distortion's type and a number of distortion's level: "iAA_BB_C.bmp". 971 experiments was conceded out by 971 observers from five countries: Finland, France, Italy, Ukraine and USA to achieve MOS which ranges from 0 to 1 with MSE 0.018 for each score. About 524340 comparisons of visual quality of distorted images or 1048680 assessments of relative visual quality in image pairs was done.

VI. Simulation Parameters

Three following three popular performance measures are used to compare performance of various metrics:

- Pearson linear correlation coefficient (PLCC)
- Spearman rank order correlation coefficient (SROCC)
- Kendall rank correlation coefficient (KRCC)

A. *Pearson linear correlation coefficient (PLCC)*

PLCC was developed by Karl Pearson. Pearson's correlation coefficient between two variables is defined as the covariance of the two variables divided by the product of their standard deviations. Statistically, it measures the linear dependence amid two variables resulting in a value having range [-1, 1] where 1 is total positive correlation and -1 is total negative correlation. Both values 1 and -1 gives extreme correlation and 0 shows that there is no correlation among the variables. If X and Y are considered as two variables and r as a Pearson correlation coefficient, then r is defined as:-

$$r = \frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2}} \quad (10)$$

B. *Spearman rank order correlation coefficient (SROCC)*

The spearman's correlation coefficient is used when both the variables are in ranked order data type called ordinal data. Let X and Y are two variables both of size n, then to determine the Spearman's correlation coefficient ρ, the n raw scores X_i, Y_i are converted to ranks x_i, y_i and ρ is given as:-

$$\rho = 1 - \frac{6\sum d_i^2}{n\left(n^2-1\right)} \quad (11)$$

where, $d_i = x_i - y_i$ denotes the difference between ranks.

The maximum value of correlation coefficient is 1 which gives perfect positive association between the ranks and the minimum value is -1 which denotes a perfect negative association between the ranks. The value of zero shows no association between the ranks.

C. Kendall rank correlation coefficient (KRCC)

Similar to Spearman rank correlation coefficient, Kendall rank correlation coefficient (also known as Kendall's tau (τ) coefficient) is aimed to evaluate the association between two ordinal (two ranked variables, not necessarily intervals) variables.

Let (x_i, y_i) be a set of views of the random variable X and Y in a manner that all values of (x_i) and (y_i) are distinctive. Any couple of ranks (x_i, y_i) and (x_j, y_j) are supposed to be *concordant* if ranks for both elements approve: both $x_i > x_j$ and $y_i > y_j$ or both $x_i < x_j$ and $y_i < y_j$. They are supposed to be *discordant*, if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. The pair is neither concordant nor discordant, if $x_i = x_j$ or $y_i = y_j$. The equation for Kendall coefficient, τ is given as:-

$$\tau = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} sgn(x_i - x_j)\, sgn(y_i - y_j)}{n(n-1)} \tag{12}$$

where, *sgn*(.) is the Signum function of its argument.

VII. Results

The simulation for obtaining the image quality scores was performed for each of the image quality metric discussed in previous sections, separately over the whole TID2013 database. Table 1 shows the performance of objective image quality metrics on TID2013 for all images in terms of accuracy, monotonicity and consistency using the correlation coefficients discussed previously.

Table 1: Performance comparison

Metric	PLCC	SROCC	KRCC
PSNR	0.566	0.653	0.482
SSIM	0.589	0.634	0.462
MSSIM	0.776	0.790	0.604
UQI	0.610	0.590	0.594
VIF	0.606	0.615	0.462
PSNRHVS	0.650	0.666	0.518
FSIM	0.822	0.810	0.636

Assessment of objective model correlations for each metric with respect to the HVS using Pearson Linear Correlation coefficient is given as:-

$$PSNR < SSIM < VIF \cong UQI \cong PSNRHVS < MSSIM < FSIM$$

This result shows that PSNR is worst predictor of image visual quality. All other six Objective Image Quality Assessment metrics are better than PSNR. Among all six algorithms, FSIM gives the best performance.

Assessment of objective model correlations for each metric with respect to the HVS using Spearman Rank Order Correlation Coefficient is given as:-

$$UQI \cong VIF < SSIM < PSNR \cong PSNRHVS < MSSIM < FSIM$$

This result indicates that UQI and VIF, both are worst interpreter of image visual fidelity. Also, PSNR and PSNRHVS gives approximately same result. Again, among all metrics the best performance is obtained in FSIM.

Assessment of objective model correlations for each metric with respect to the HVS using Kendall Rank Correlation Coefficient is given as:-

$$UQI < VIF < SSIM < PSNR \cong PSNRHVS < MSSIM < FSIM$$

This result shows that UQI is the worst predictor of visual quality. Rest of the metrics give better performance. Again, FSIM gives best performance among all the seven algorithms.

VIII. Conclusion

In this paper, overall performance of various objective image quality assessment metrics is compared via simulations using publicly available image database with a wide range of distortion types. Seven commonly used and publicly-available quality assessment methods are studied. The results obtained shows that different metrics perform differently with respect to different correlation method. PSNR gives poorest result with respect to PLCC whereas, UQI and VIF gives worst result with respect to SROCC and KRCC method. But in all correlation methods, FSIM gives the best performance among seven metrics studied in this paper.

Subjective quality assessment methods cannot be used in real-time applications. So Objective quality assessment methods are widely used in recent years. But only more precise and efficient method prove their applicability in real-time systems.

References

[1] T. Seemann and others, Digital image processing using local segmentation, Monash University, 2003.
[2] Z. Wang, A. C. Bovik and L. Lu, "Why is image quality assessment so difficult?," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, 2002.
[3] Y. A. Al-Najjar, "Comparison of Image Quality Assessment: PSNR, HVS, SSIM, UIQI," *International Journal of Scientific \& Engineering Research*, vol. 3, no. 8, pp. 2229-5518, 2012.
[4] R. F. ÖLGÜN, "Evaluation Of Visual Quality Metrics," 2011.
[5] A. Lahouhou, E. Viennet and A. Beghdadi, "Combining and selecting indicators for image quality assesment," in *Information Technology Interfaces, 2009. ITI'09. Proceedings of the ITI 2009 31st International Conference on*, 2009.
[6] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600-612, 2004.
[7] Z.-S. Xiao, "A Multi-scale Structure SIMilarity metric for image fusion qulity assessment," in *Wavelet Analysis and Pattern Recognition (ICWAPR), 2011 International Conference on*, 2011.
[8] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, 2003.

[9] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430-444, 2006.
[10] L. Zhang, D. Zhang and X. Mou, "FSIM: a feature similarity index for image quality assessment," *Image Processing, IEEE Transactions on*, vol. 20, no. 8, pp. 2378-2386, 2011.
[11] Z. Wang and A. C. Bovik, "A universal image quality index," *Signal Processing Letters, IEEE*, vol. 9, no. 3, pp. 81-84, 2002.
[12] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti and M. Carli, "New full-reference quality metrics based on HVS," in *CD-ROM proceedings of the second international workshop on video processing and quality metrics, Scottsdale, USA*, 2006.
[13] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti and others, "Color image database TID2013: Peculiarities and preliminary results," in *Visual Information Processing (EUVIP), 2013 4th European Workshop on*, 2013.

Raman Gupta received his B.E (Electronics and Communication) degree from Government College of Engineering and Technology, Jammu, India, in the year 2011. He is pursuing M.Tech. degree in Electronics and Communication at Punjabi University, Patiala, Punjab, India.

Er. Dipti Bansal is working as Assistant Professor at the Department of Electronics and Communication Engineering, Punjabi University, Patiala, Punjab, India.

Dr. Charanjit Singh is working as Assistant Professor at the Department of Electronics and Communication Engineering, Punjabi University, Patiala, Punjab, India.