Prediction and Measurement of High Quality in Still Image Coding

Niranjan Avadhanam and **V Ralph Algazi**

Center for Image Processing and Integrated Computing (**CIPIC**)
University of California, Davis

ABSTRACT

In image coding applications, quantitative quality metrics which approximate the perceived quality of an image can be used for evaluating the performance of coders, designing new coders and optimizing existing coders. In this paper, we consider images of high quality levels for which most of the errors are in the threshold region of perception, and perceptible errors are sparsely distributed. Two different methodologies are used; first, we use an objective Picture Quality Scale (PQS) which combines partial measures, denoted distortion factors, of random as well as structured perceived distortions due to coding. We also consider an alternative approach, applicable at high quality, that is based on a multi-channel vision model, the Visible Differences Predictor (VDP) proposed by Scott Daly. The VDP produces a probability detection map identifying regions in which errors are sub-threshold, threshold and supra-threshold. For the PQS, since distortion factors due to structured errors and errors in the vicinity of high contrast image transitions are most important at high quality, these two factors are analyzed to compute spatial distortion maps of their contributions. This paper compares the spatial distortion maps produced by both methods for high quality still images to experimental observations. Global values for the quality have been obtained by integrating the factor images obtained using PQS and yield a correlation of 0.9 with mean opinion of score values for JPEG, Subband and Wavelet coded images.

Keywords : image quality, perceived distortions, objective quality, multichannel vision models, visual quality assessment, threshold of perception.

1 INTRODUCTION

Evaluating the quality of a picture is an indispensable part of visual communications. Quality evaluation can be done by two methods : subjective testing or by applying a quantitative objective measure. Subjective testing is time consuming and suffers from variability due to observers and to the testing conditions. Hence, objective measures which produce reproducible results and reflect the perceived quality of images, are attractive. Since the human viewer is the final judge of image quality, the objective measure should approximate as closely as possible, the human visual quality assessment.

Objective measures of quality, classified as either space domain models or as frequency domain models[1] have attracted much attention in the past two decades.[2–4] Some recent work has addressed specific distortions in coded images, primarily blocking artifacts.[5] For perception based models, in recent work, many researchers use the multiple channel narrow band model of vision which analyzes images by filtering with frequency and orientation selective filters.[6–9]

1.1 Prediction of quality

By prediction of quality, we mean the algorithmic determination of the location and severity of perceived errors. The PQS approach is based on classifying and quantifying the actual distortions produced by coders and defining a set of partial distortion measures.[10,11] These local measures are combined to obtain spatial factor images which show the magnitude and locations of the errors. The VDP produces a map of the probabilities of detecting the distortions, classifying them as sub-threshold, threshold and supra-threshold.

We shall consider the prediction of visible distortions using these two approaches. The vision model attempts to predict in a universal way, the distortions perceived by humans. The spatial distortion model identifies, and then quantifies the practical distortions coders produce and attempts to predict on a local scale, what the important distortions are, and where they occur in the image. Preserving the distortion phase information or using distortion maps allows the user to identify the location of distortions in evaluating the quality of a coding algorithm. The maps can also be used to produce Just Noticeable Differences (JND) or Minimum Noticeable Differences (MND) maps. These could be used as cues for a coding algorithm for a variable bit rate, constant quality or constant bit rate, variable quality performance.[12] This information may also be of value to the design of any processing algorithm in that the designer can assess the type of visible distortions and their location.

1.2 Measurement of quality

The second issue of interest in this paper is the measurement of quality. Although information is obtained from the distortion maps, there is also a need for composing the distortion maps into a single number to reflect the quality of the image. For the VDP, this has not been attempted, although a preliminary step would be to accumulate the errors in a few severely impaired portions of the image.[13] The PQS however, was originally developed to measure the global quality of images by matching the PQS values to the Mean Opinion of Scores (MOS[1]) values of a set of test images. From the factor images, the localized distortions are summed to obtain global values for distortion factors, a linear combination of which gives the overall PQS value.

At high quality, we normally deal with images which have distortions that are near threshold so that the visual system can be assumed to behave linearly for incremental changes. The distribution of the errors weighted with respect to their visibility is an indicator of the visual quality. Most of the images in this study have a MOS of more than 3.5 on a five point impairment scale.[1]

This paper is divided into four sections. In section 1, a brief discussion and illustration of the PQS is given. Section 2 deals with a discussion on the VDP. In section 3, we describe the application of these methods to high quality still images which have been coded using JPEG, subband and wavelet coders. In section 4, we discuss the experiments done and the results. Finally, we present the conclusions. In the terminology used in this paper, factor images denote the map of the localized distortions. Distortion maps are obtained from the factor images by using a perceptual threshold and the global distortion factors are obtained by summing each factor image to get a single number.

2 THE PQS ALGORITHM

The PQS for the evaluation of coded images is obtained by considering both perceptual properties and the actual image disturbances or errors that can be observed in the coding of real images.[11,10,14] First, the image signal is transformed into one which is proportional to the visual perception of luminance according to the Weber-Fechner's Law and then the frequency weighted errors $e_w(m,n)$ are obtained. Perceived image disturbances are described and the corresponding objective quality factors which quantify each image degradation are presented. The perceived disturbances are D_i and the defined numerical measures of them are the corresponding distortion factor F_i. The distortion factors F_i are defined as numerical functions of $e_w(m,n)$. The original PQS made use five such factors, the first three being two factors for random errors and the third for blocking. But here we use only the last two factors which correspond to correlated errors, F_4, and errors at high contrast transitions, F_5 that are relevant for high quality images. We assume that the MOS is to be approximated by a linear combination of the distortion factors. Since the distortion factors are functions of the difference between the original and the reconstructed encoded image, PQS will be a measure of degradation from an original, rather than an absolute image quality measure.

2.1 Distortion Factor F_4 = General Correlated Errors

Image structures lead to structured coding errors in their vicinity. These errors with strong correlation are much more perceptible than random errors. Based on a detailed experimental study, F_4 is defined by using correlation coefficients, so that[15]:

$$F = ({R_x}^2 + {R_y}^2)^{\frac{1}{2}} \tag{1}$$

Where R_x and R_y measure the correlation and thus the structure of errors horizontally and vertically and F_4 will be zero if errors are uncorrelated. The factor image is obtained by performing this computation locally for lags upto 8 pixels. These local computations give the factor image, which is then summed to obtain the global value for the distortion factor F_4.

2.2 F_5 = Random errors in the vicinity of high contrast image transitions

Large coding errors occur at, and in the vicinity of high contrast transitions. Even though these visual disturbances will be masked by the image activity, they will still be perceptible, especially when they extend into flat portions of the image adjacent to the transitions. If we define the local activity function as

$$V_x(r,s) = \left| \frac{I(r-1,s) \;-\; I(r+1,s)}{2} \right| \tag{2}$$

then

$$S_x(r,s) = e^{\{-0.04 V_x(r,s)\}} \tag{3}$$

gives the masking factor based on the local activity. The error after masking is

$$e_x = \left| \; e_w(r+j,\; s)\; S_x(r+j,\; s) \; \right| \text{, for all } j \in J \tag{4}$$

$S_x(r,s)$ is the masking factor, which can be substantially less than unity if $V_x(r,s)$ is large. e_y is similarly defined. J is a neighborhood of 2 ℓ pixels at the vicinity of significant edge points in the image, where the edge points are detected by using $[3 \times 3]$ Kirsch operator. This neighborhood 2 ℓ depends on the viewing distance and for the present work, ℓ is chosen to be 7. We obtain the value of F_5 as

$$F_5 = ({e_x}^2 + {e_y}^2)^{\frac{1}{2}} \tag{5}$$

F_5 thus measures, with a suitable weight to account for visual masking,[16] these large localized errors. The factor image is obtained by performing these computations on a local scale and the global distortion factor F_5 is obtained by summing these localized errors as an expectation taken only over the edge points.

The global value for PQS is given by the linear combination of the factors F_j as

$$PQS = b_o + \sum_{j=1}^{J} bj \;\cdot Fj \tag{6}$$

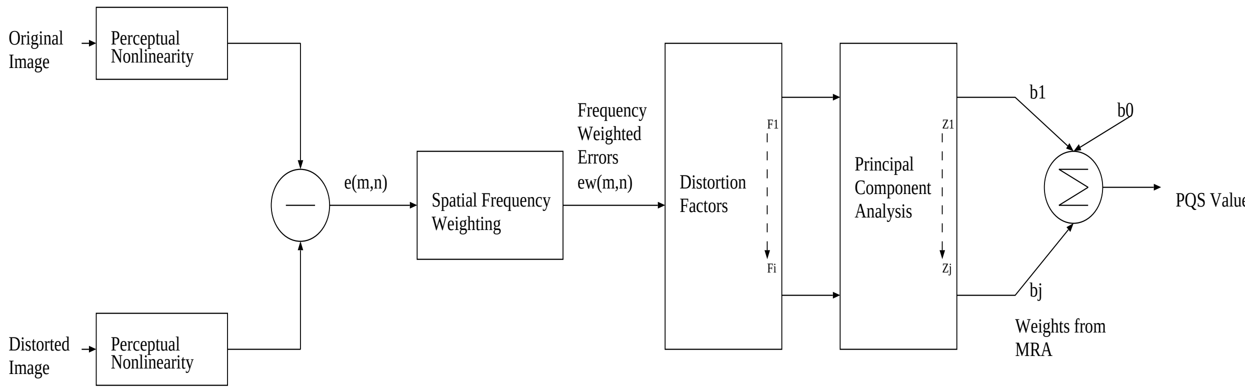

Figure 1: Overall PQS

where b_O is a fixed parameter and b_j are the partial regression coefficients which are computed by Multiple Regression Analysis (MRA)[17] between the factors and the MOS values of observers obtained by experimental quality assessment tests. Figure 1 summarizes the steps used in the construction of PQS.

We determine the goodness of the approximation between the obtained PQS and the MOS. The correlation coefficient R is 0.92, which is a great improvement when compared to the correlation of R=0.47 of the conventional WMSE scale which is calculated by using F_1 only.

3 THE VDP ALGORITHM

The VDP algorithm proposed by Scott Daly is a multichannel model which takes an image processing approach to the quality prediction. The overall model is implemented as a cascade of submodels to account for known sensitivity variations of the visual system. The sensitivity variations are due to light level, spatial frequency and content in the image. Figure 2 gives an overall view of the algorithm. Complete details are given in references.[7,18] A number of psychophysical tests have been performed to test the validity of this algorithm.[19,20]

A shift invariant nonlinearity models the light adaptive property of the retina. The variation with respect to spatial frequency is modeled by a Contrast Sensitivity Function (CSF). The CSF is a function of spatial frequency, orientation, light adaptation level, image size and lens accomodation. A fixed viewing distance of four times the picture height has been assumed for the present work.

The sensitivity due to image content is modeled by the detection process which consists of a number of frequency and orientation tuned channels, a nonlinearity to account for masking, a psychometric function to map the contrast errors into probabilities and finally a summation stage to combine all the probabilities into a single map of probabilities of detection of the errors as a function of their location in the image.

The decomposition into multiple spatial frequency and orientation tuned channels is achieved by a cascade of frequency selective filters (denoted as dom filters[7]) and orientation selective fan filters based on the Cortex Transform.[21] The dom filters have octave bandwidths and are symmetric on a log frequency axis as shown in figure 3. The fan filters have a tuning bandwidth of 30 degrees. This selectivity yields specific frequency and orientation tuned bands called cortex bands. The present implementation has 5 doms and 6 fans yielding a total

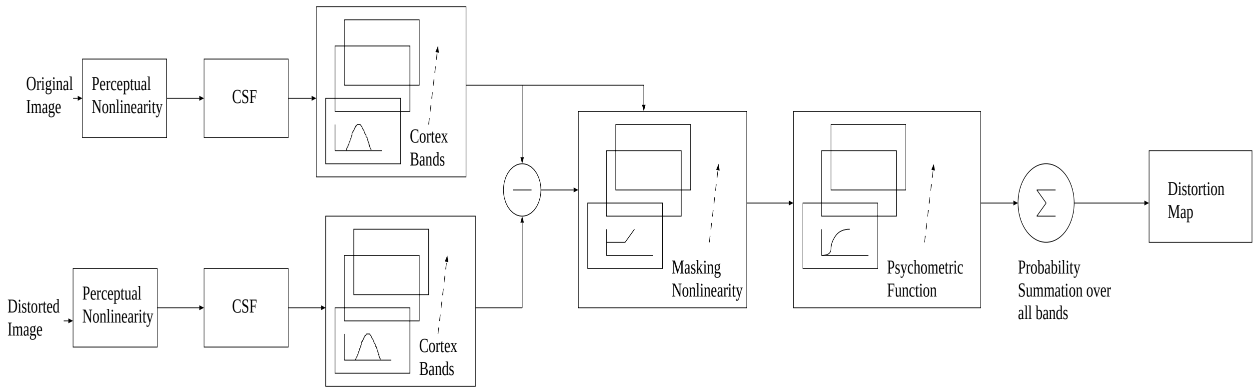

Figure 2: Overall VDP

of 31 cortex bands including the baseband.

The masking effect due to the image activity is evaluated for each cortex band and results in a threshold elevation due to masking calculated as a nonlinear function of the normalized mask contrast m_n, given by

$$T_e = (1 + (k_1(k_2 m_n)^s)^b)^{\frac{1}{b}} \tag{7}$$

Where s is between 0.7 and 1, $k_1=6^{\frac{-7}{3}}$, $k_2=6^{\frac{10}{3}}$ and b=4. The choice of these parameter values is explained in the reference.[7] The contrast difference of the errors for each location in a band is mapped through a psychometric function of the form

$$P(c) = 1 - e^{-(\frac{c}{\alpha})^\beta} \tag{8}$$

where c is the contrast of the error, α the contrast threshold, and β the slope of the psychometric function. This yields a probability of detection map for each band. Since the channels are independent, an error above threshold in any of the cortex bands would be perceivable. Hence the probability maps for all the 31 bands are combined as below to give a single map of the probability of detection of the errors as a function of location in the image.

$$P_t(i,j) = \prod_{k=1..5, l=1..6} (1 - P_{k,l}(i,j)) \tag{9}$$

4 DISTORTION MAPS TO PREDICT QUALITY

There are several advantages gained by the use of distortion maps for quality prediction and evaluation. If the distortion map corresponds for example, to a threshold map, it may be used to drive an image compression algorithm to limit the distortions so that perceptually lossless compression can be achieved.[12,8] When the target bitrate is no longer compatible with this level of quality, the maps can still be used to obtain perceptually lossy but optimum performance.[12] Distortion maps could also be used by the algorithm designer to quickly verify what errors are produced and where they occur in the image.

In the case of methods quantitative measures like the PQS, the evaluation and verification of the coding algorithm is more rigorous (especially at high quality) when localized errors are considered. When an attempt is made to integrate the maps into a single measurement, some points have to be considered. Image quality expressed as a single number, is a summation over the whole picture and does not reflect the way in which the errors are distributed. It might also not reflect the way the subject evaluates a picture because the subject typically tends

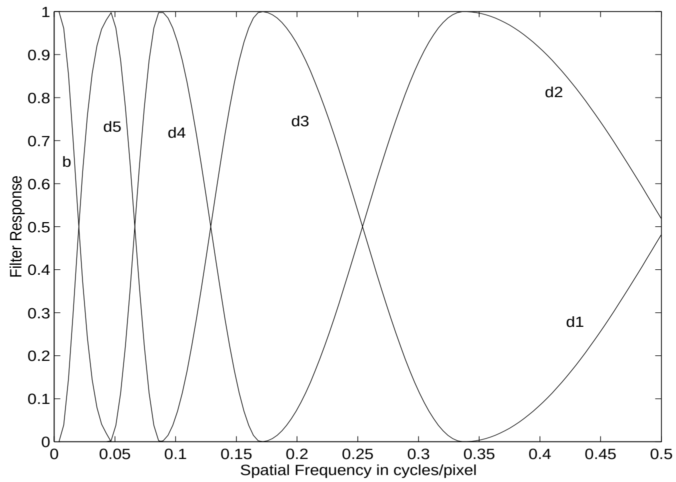

Figure 3: Dom Filters

to concentrate on a few parts of the image where the most severe degradations occur and where the area of the impaired region is large.[13]

In the PQS system, we consider the spatial distortion maps, which are the local contributions to the distortion factors that make up PQS, as spatial indicators of perceptible distortions. At high quality, perceptible errors are associated with three significant features: high contrast transitions or edges, high detail regions over a flat background and diagonal structures. The localized maps for F_4 and F_5 that is, the factor images, capture these distortions. Since we are able to localize the portions and features of images that result in significant values of the distortion factors, we can produce thresholded maps for F_4 and F_5 factor images and a combined distortion map reflecting their perceptual weights. The threshold used to obtain the distortion map from the factor image was obtained empirically after studying different kinds of distortions. The distortion maps are now essentially binary images, showing the locations of the visible errors and ignoring their magnitude.

For the VDP, once the probability summation is done, we obtain a map of probabilities of detecting errors as a function of location. For the present work, the sign of the error is ignored; hence the probabilities range from 0 to 1. The gray level distortion maps are displayed as free-field difference maps.[7] The gray level value of the map is given by

$$I(i,j) = P_t(i,j) * (\frac{255-0}{2}) + (\frac{255+0}{2}) \tag{10}$$

Pixels which are lighter than 128 indicate the visible errors in the threshold region and above and values of 255 correspond to fully detected errors.

5 EXPERIMENTAL RESULTS AND DISCUSSION

A number of experiments were performed to test the results of the prediction models on monochrome images compressed by JPEG, Subband and Wavelet coders. The results were informally tested on a HP workstation monitor with a gamma of 2.2. All the distortion maps were further verified on a Super Match monitor connected to a Macintosh Power PC whose display was calibrated accurately. The display had a gamma of 2 and was set to accomodate a dynamic range of 100. Ambient light levels were reduced to a minimum.

The images used were of size either 512*512 or 256*256 and the viewing distance was fixed to four times the picture height. The experimental results show that the factor images produced by F_4 correspond well to the locations of the images where structured errors occur. The factor images produced by F_5 corresponds to errors near high transition regions after taking into account the masking which occurs there. The combined binary distortion map primarily identifies the portions where structured errors and 'unmasked' errors are visible. The VDP gray level distortion map predicted locations of subthreshold, threshold and suprathreshold errors in fairly good correspondence to perceived distortions.

As expected, the binary distortion maps produced by the VDP and PQS resemble each other and we can say that PQS predicts the structured errors and unmasked errors at the same locations the VDP indicates errors to be visually relevant. Figures 4 and 5 show the distortion maps produced for the image 'Lenna' JPEG coded (using the Independent JPEG group's free software) with a quality setting of 40 which is a reasonably high quality level. The distortion map shown in figure 4 is a binary image, specifying the locations of visible errors, independently of their magnitude and is obtained from thresholding the factor images. The map using the VDP in figure 5 shows only the errors which the algorithm predicts to be clearly suprathreshold.

From the experimental observations, we noticed that the VDP underestimates some errors in the light portions of the image and this may be an effect of the perceptual grayscale nonlinearity. Because some filters in the VDP filter bank have large extents in the spatial domain, filter ringing effects can be seen in the maps. The learning slope for the masking curve was set to 1 for all the cortex bands.[7] In the PQS, the masking effect is quantified by multiplying the errors near significant edges by an exponential function of the activity in the vicinity of the errors. This masking function does not account well for visual masking. Hence, the PQS map for F_5 tends to overestimate the errors present near significant edges. Secondly, the thresholds used to get the distortion maps were obtained empirically based on observing a number of images and we combine the maps for F_4 and F_5 by an 'OR' operation. More refined ways of combining these maps are being studied and should give better results. Still, we can see that the maps appear reasonably similar to each other and are resonably successful in predicting the locations of visible errors.

To measure the quality of the images, the PQS factor images were integrated over the whole image into two numbers corresponding to F_4 and F_5. Coefficients obtained by performing a MRA with the MOS values gave resultant global values for PQS which had a correlation of 0.9 with the MOS values. The test set consisted of a set of 28 images which were obtained by using JPEG, Wavelet and Subband coders on five different images.

6 CONCLUSIONS

We have evaluated two methods for the prediction of visible errors produced by high quality image compression. The distortion maps produced by the vision model and the spatial distortion model agree with each other and can be used for predicting visible coding errors. The PQS maps have been integrated into a number which approximates the MOS value of the images well. Additional work is underway to refine the methods used to obtain the prediction maps and to integrate the maps into a single measurement reflecting the visual quality. As mentioned before, there are many applications for the distortion maps and the global distortion metric and both the PQS and the VDP approaches can provide some of the necessary methodology and quantitative tools for these applications.

7 ACKNOWLEDGMENTS

We acknowledge the support of this research by the UC MICRO program, Pacific Bell, Lockheed Martin and Hewlett Packard.

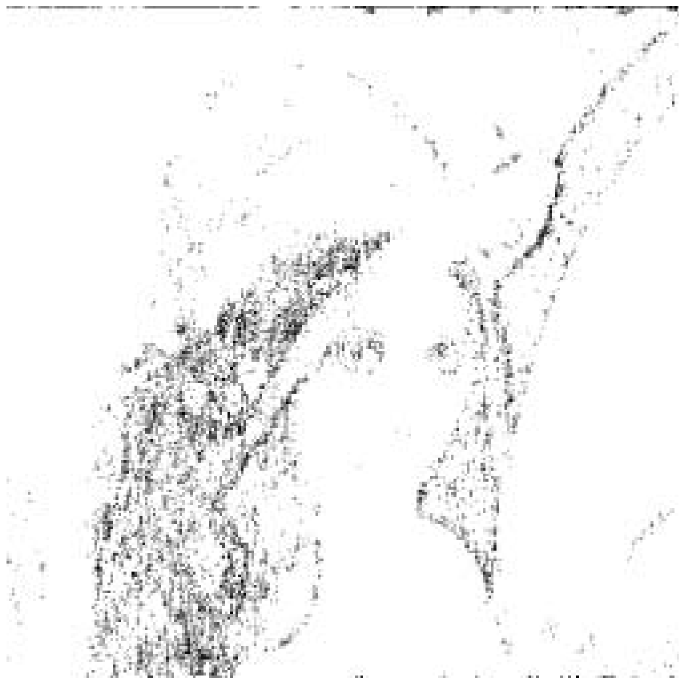

Figure 4: Combined PQS distortion map

Figure 5: VDP distortion map

8 REFERENCES

[1] A.N. Netravali and B.G. Haskell, "Digital Pictures," *Charles Griffin*, 1988.

[2] F. X. J. Lukas and Z. L. Budrikis, "Picture Quality Prediction Based on a Visual Model," *IEEE Transactions on Communication*, Vol. COM-30, No. 8, pp. 1679-1692, July 1982.

[3] J.O. Limb, "Distortion criteria of the human viewer," *IEEE Transactions on System, Man, and Cybernetics*, SMC-9(12), pp.778-793, December 1979.

[4] D.J. Sakrison, "On the role of the observer and a distortion measure in image transmission," *IEEE Transactions on Communication*, Vol. COM-25, No. 11, November 1977.

[5] S. A. Karunasekera and N. G. Kingsbury, "A distortion measure for blocking artifacts in images based on human visual sensitivity", *IEEE Transactions on Image Processing*, Vol. 4, No. 6, pp. 713-724, June 1995.

[6] C. Zetzsche and G. Hauske, "Multiple channel model for the prediction of subjective image quality," *Society of Photo Instrumentation Engineering Proceedings*, SPIE Vol. 1077, pp. 209-216, 1989.

[7] S.Daly, "The visible differences predictor : an algorithm for the assessment of image fidility," *Society of Photo Instrumentation Engineering Proceedings*, SPIE Vol. 1666, pp. 2-15, 1992.

[8] K. S. Prashant, V. J. Mathews and P. J. Hahn, "A new model of perceptual threshold functions for application in image compression systems", *Data Compression Conference*, DCC 95, pp. 371-380, 1995.

[9] P. C. Teo and D. J. Heeger, "Perceptual image distortion", *International Conference on Image Processing*, ICIP Vol. 2, pp. 982-986, 1994.

[10] V. R. Algazi, Y. Kato, M. Miyahara and K. Kotani, "Comparision of image coding techniques with a picture quality scale,"*Society of Photo Instrumentation Engineering Proceedings*, SPIE Vol. 1771, pp. 396-405, 1992.

[11] M. Miyahara, "Quality assessments for visual service," *IEEE communications Magazine*, 26, 10, pp. 51-60, October 1988.

[12] N. Jayant, J. Johnston and R. Safranek, "Perceptual coding of images,"*Society of Photo Instrumentation Engineering Proceedings*, SPIE Vol. 1913, pp. 168-178, 1993.

[13] S. Ohtsuka, M. Kosugi, "Quality evaluation of locally impaired pictures," *Society for Information Display Digest*, SID 90 Digest, pp. 33-36, 1990.

[14] V. R. Algazi, H. Ohira, K. Kotani and M. Miyahara, "Important distortion factors in the encoding of very high quality images,"*Society of Photo Instrumentation Engineering Proceedings*, SPIE Vol. 2298, pp. 2-12, 1994.

[15] Y. Horita and M. Miyahara, "Image coding and quality estimation in uniform perceptual space," *IECE Tech. Report*, IE87-115, pp. 25-32, January 1987.

[16] J.O. Limb, "On the design of quantizer for DPCM coder: A functional relationship between visibility, probability, and masking," *IEEE Transactions on Communications*, Vol. VOM-26, No. 5, pp. 573-578, May 1978.

[17] M.G. Kendall, "Multivariate Analysis," *Charles Griffin*, 1975.

[18] S.Daly, "The visible differences predictor : an algorithm for the assessment of image fidility," in "Digital Images and Human Vision",*MIT Press*, pp. 179-206, 1993.

[19] S.Daly, "A visual model for optimizing the design of image processing algorithms," *International Conference on Image Processing*, ICIP Part 3, pp. 16-20, 1994.

[20] S.Daly, "Quantitative performance assessment of an algorithm for the determination of image fidelity," *Society for Information Display Digest*, SID 93 Digest, pp. 317-320, 1993.

[21] A.B.Watson, "The cortex transform : Rapid computation of simulated neural images," *Computer Vision Graphics and Image Processing*, CVGIP Vol. 39, pp. 311-327, 1987.