Title
Objective Picture Quality Scale (PQS) For Image Coding

Authors
Miyahara, M.
Kotani, K.
Algazi, Ralph

Publication Date
1996

Peer reviewed

(Submitted to IEEE Transaction on Communications)

Objective Picture Quality Scale (PQS) For Image Coding

M. Miyahara*, K. Kotani*, and V.R. Algazi**

***Japan Advanced Institute of Science and Technology (JAIST), East**
**** CIPIC, Center for Image Processing and Integrated Computing,**
University of California, Davis[1]

Abstract

A new methodology for the determination of an objective metric for still image coding is reported. This methodology is applied to obtain a Picture Quality Scale (PQS) for the coding of achromatic images over the full range of image quality defined by the subjective Mean Opinion Score (MOS). This Picture Quality Scale takes into account the properties of visual perception for both global features and localized disturbances. PQS closely approximates the MOS, with a correlation coefficient of more than 0.92, as compared to 0.57 obtained using the conventional WMSE. Extensions and applications of the methodology and of the resulting metric are discussed.

1 Introduction

The evaluation of picture quality is indispensable in image coding. Subjective assessment tests are widely used to evaluate the picture quality of coded images [1, 2, 3]. However, careful subjective assessments of quality are experimentally difficult and lengthy, and the results obtained may vary depending on the test conditions. Further, subjective assessments provide no constructive methods for performance improvement, and are difficult to use as part of the design process.

Objective measures of picture quality would not only alleviate the difficulties described above, but would also help expand the field of image coding. This expansion would result from the systematic determination of objective measures for the comparison of coded images, and also from the possibility of successive adjustments to improve or optimize the picture quality for a desired quality of service [4]. The objective simulation of performance both with respect to bit rate and image quality would also lead to a more systematic design of image coders.

It is important that an objective scale mirror the perceived image quality. For instance, simple distortion scales, such as the signal to noise ratio (PSNR), or even the weighted mean square error (WMSE) [5] are good distortion indicators for random errors, but not for structured or correlated errors. But such structured errors are

[1]Research supported in part by travel grants from NSF in the USA and JSPS in Japan, and also by the UC MICRO program, Lockheed, Pacific Bell, Hewlett Packard, Mitsubishi Electric and NEC.

prevalent in image coders, and degrade local features and perceived quality much more than do random errors [6]. Hence, PSNR and WMSE alone are not suitable objective scales to evaluate compressed images. There have been many studies of the construction of objective scales which represent properties of the human observer [7, 8, 9]. We note among the early studies, the work of D.J. Sakrison, who proposed a picture quality scale and gave an integrated view, for image coding applications, of known perceptual properties of vision [10]. Other models and applications of perception to coding have been been reported [11, 12, 13] An additional discussion of more recent work on perceptual models and their application to coding performance evaluation is given later in the paper.

This paper proposes a new methodology for the determination of objective quality metrics, and applies it to obtain a Picture Quality Scale (PQS) for the evaluation of coded achromatic still images. The approach is based on the perceptual properties of human vision and extensive engineering experience with the observation of actual image disturbances resulting from the image coding.

The properties of perception suggest the transformation of the images and coding errors into perceptually relevant signals. First, we transform the image signal into one which is proportional to the visual perception of luminance using the Weber-Fechner's Law and the contrast sensitivity for achromatic images. Secondly, we apply spatial frequency weighting to the errors. Third, we describe perceived image disturbances and the corresponding objective quality factors which quantify each image degradation. In this step, we include visual masking where it is relevant to the perception of image degradations. Fourth, we describe the experimental method for obtaining PQS based on these distortion factors, and we determine the goodness of the approximation between the obtained PQS and the Mean Opinion Score (MOS). In the discussion section, limitations, extensions and refinements of the PQS methodology are considered. Application and implications of PQS to coder comparison and to a systematic image coding design are finally considered briefly [14, 15, 16].

2 Construction of a Picture Quality Scale (PQS)

The PQS methodology is illustrated in Figure 1. Given the original image $i(m,n)$ and a distorted, compressed image $\hat{i}(m,n)$, we compute local distortion maps $\{f_i(m,n)\}$, from which the distortion factors $\{F_i\}$ are computed. We then use regression methods to combine these factors into a single number representative of the quality of a given image.

In the sections that follow, we discuss in detail the various components of this framework, starting with the computation of contrast adjusted error images, and concluding with explicit formulas for the computation of the PQS factors and then PQS itself.

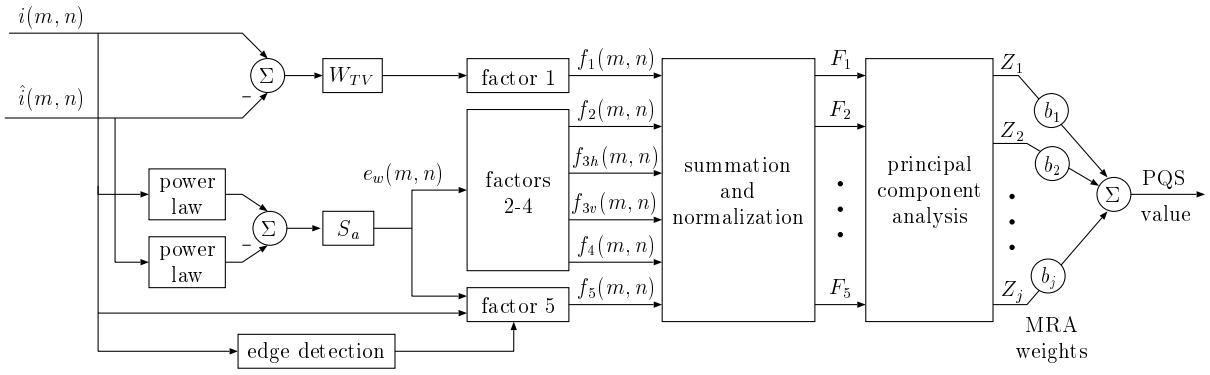

Figure 1: The construction of PQS.

2.1 Luminance Coding Error

To provide a more uniform perceptual scale, we transform the images using

$$x(m,n) = k \cdot i(m,n)^{1/2.2}, \tag{1}$$

which closely approximates Weber-Fechner's Law for contrast sensitivity. The contrast adjusted error image is then computed as

$$e(m,n) = x(m,n) - \hat{x}(m,n) \tag{2}$$

where $\hat{x}(m,n)$ is the contrast adjusted version of $\hat{i}(m,n)$. Most of the distortion factors are defined as functions of $e(m,n)$.

2.2 Spatial Frequency Weighting of Errors

The contrast sensitivity function of vision suggests a spatial or spatial frequency distortion weighting.

Based on a measured contrast sensitivity function, the spatial frequency response is modeled approximately by

$$S(\omega) = 1.5e^{-\sigma^2\omega^2/2} - e^{-2\sigma^2\omega^2}, \tag{3}$$

where

$$\sigma = 2,\ \omega = \frac{2\pi f}{60},\ f = \sqrt{u^2+v^2}, \tag{4}$$

and u and v are the horizontal and vertical spatial frequencies, respectively, in cycles per degree.

At higher spatial frequencies, the frequency response is anisotropic so that a better model [17] is given by

$$S_a(u,v) = S(\omega)O(\omega,\theta), \tag{5}$$

with

$$O(\omega,\theta) = \frac{1 + e^{\beta(w-w_o)}\cos^4 2\theta}{1 + e^{\beta(w-w_o)}}, \tag{6}$$

where $\theta = \tan^{-1}(v/u)$ is the angle with respect to the horizontal axis,

$$\beta = 8,\ f_o = 11.13 \text{ cycle/degree}, \tag{7}$$

and $O(\omega,\theta)$ blends in a $\cos^4 2\theta$ anisotropy, fairly quickly, for frequencies $f > f_o$. The frequency weighted error $e_w(m,n)$, then, is just the contrast adjusted error, filtered with $S_a(u,v)$.

In our presentation, perceived disturbances are first described verbally, for example, as random or structured errors. Next, these disturbances are quantified locally, resulting in distortion maps or factor images $\{f_i(m,n)\}$. A single numerical value, or distortion factor F_i, is then computed from each factor image. We now describe the perceived disturbances and define numerical measures of the corresponding distortion factors $\{F_i\}$. We make use of five such factors and then analyze their relative importance.

2.3 Random Errors and Disturbances

All coding techniques will produce random errors. The perceived random disturbances will be in the form of incremental noise in slowly varying regions of the image and, generally, will be not perceived in active areas of the image. We use an integral square measure to compute distortion factors F_1 and F_2.

Because the CCIR has adopted a standard to quantify the effect of noise, we follow that standard in defining F_1 [5].

2.3.1 Distortion Factor F_1

The CCIR 567-1 television noise weighting standard does not take into account Weber's law. Thus, we define

$$e_i(m,n) = i(m,n) - \hat{i}(m,n), \tag{8}$$

and compute the first factor image,

$$f_1(m,n) = [e_i(m,n) * w_{TV}(m,n)]^2, \tag{9}$$

where the frequency weighting defined by CCIR 567-1,

$$W_{TV}(f) = \frac{1}{1 + (f/f_c)^2},\ f = \sqrt{u^2 + v^2}, \tag{10}$$

with a 3 dB cutoff frequency $f_c = 5.56$ cycles/degree at a viewing distance of 4 times picture height ($4H$), and $*$ represents convolution.

The distortion factor F_1 is then computed as

$$F_1 = \frac{\sum_{m,n} f_1(m,n)}{\sum_{m,n} i^2(m,n)} \tag{11}$$

where the sums are computed over all pixels in the $M \times N$ images.

2.3.2 Distortion Factor F_2

Distortion factor F_2 includes a more complete single channel model of visual perception. A correction for Weber's Law (1) and the frequency weighting factor of (5) are now used. In addition, F_2 ignores values of $e_w(m,n)$ which are below a perceptual threshold T. Thus,

$$f_2(m,n) = I_T(m,n)[e_w(m,n) * s_a(m,n)]^2, \tag{12}$$

and

$$F_2 = \frac{\sum_{m,n} f_2(m,n)}{\sum_{m,n} \hat{i}^2(m,n)}, \tag{13}$$

where $I_T(m,n)$ is an indicator function for a perceptual threshold of visibility, and $T = 1$.

2.4 Structured and Localized Errors and Disturbances

Because the perception of structured patterns is more acute, and since structured and correlated errors are prevalent in coded images, we now define three additional factors to evaluate the contribution of correlated errors.

2.4.1 Distortion Factor F_3 (End of Block Disturbances)

We are specially sensitive to linear features in images and therefore to such features in errors as well. Such structured disturbances are quite apparent in most coders. These disturbances occur in particular at the end of blocks for transform coders, and are due to error discontinuities.

We define the distortion factor F_3 as a function of two factor images, one each for the horizontal and vertical block error discontinuities. Thus,

$$f_{3h}(m,n) = I_h(m,n)\Delta_h^2(m,n) \tag{14}$$

where

$$\Delta_h(m,n) = e_w(m,n) - e_w(m,n+1) \tag{15}$$

and $I_h(m,n)$ is an indicator function which selects only those differences which span horizontal block boundaries. Now,

$$F_{3h} = \frac{1}{N_h} \sum_{m,n} f_{3h}(m,n), \tag{16}$$

where $N_h = \sum_{m,n} I_h(m,n)$ is the number of pixels selected by the corresponding indicator function, and

$$F_3 = \sqrt{F_{3h}^2 + F_{3v}^2}, \tag{17}$$

with F_{3v} defined similarly at vertical block boundaries. Note that more elaborate models for end of block errors have been proposed recently [18].

2.4.2 Distortion Factor F_4 (Correlated Errors)

Even if they do not occur at the end of a block, image features and textures with strong spatial correlation are much more perceptible than random noise. In order to evaluate structured errors, we make use of their local spatial correlation. The distortion factor F_4 is thus defined as a summation over the entire image of local error correlations.

We compute locally the factor image

$$f_4(m,n) = \sum_{(k,l)\in W} |r(m,n,k,l)|^{0.25}, \tag{18}$$

where the local correlation

$$\begin{aligned} r(m,n,k,l) \;=\; & \frac{1}{n-1}\Big[\sum e_w(i,j)e_w(i+k,j+l) \\ & -\frac{1}{n}\sum e_w(i,j)\sum e_w(i+k,j+l)\Big], \end{aligned} \tag{19}$$

and the sums are computed over the set of pixels where (i,j) and $(i+k,j+l)$ both lie in a 5×5 window centered at (m,n) and W is the set of lags to include in the computation. We include all unique lags with $|k|,|l| \le 2$, except for $(0,0)$, which is the error variance. Due to symmetry, only 12 lags are included in the sum. Note that the 0.25 exponent is used to deemphasize the relative magnitude of the errors, as compared to their correlation or structure. Now, the distortion factor

$$F_4 = \frac{1}{MN} \sum_{m,n} f_4(m,n). \tag{20}$$

2.4.3 Distortion Factor F_5 (Errors in the vicinity of high contrast image transitions)

Two psychophysical effects affect the perception of errors in the vicinity of high contrast transitions: visual masking, which refers to the reduced visibility of disturbances

in active areas, and enhanced visibility of misalignments, even when they are quite small. Here, we account only for the masking effect. The second effect, denoted edge busyness is prevalent for very low bit rate DPCM coders but is not a major effect for other commonly used coding techniques, such as those considered in this work.

Although visual masking reduces the visibility of impairments in the vicinity of transitions, coding techniques will introduce large errors and major visual disturbances in the same areas. Thus, even though these disturbances are masked, they will still be most important. The distortion factor F_5 measures all disturbances in the vicinity of high contrast transitions. In contrast to the other factors, F_5 is based on an analysis of the original image, as well as the contrast enhanced error image $e_w(m,n)$.

A horizontal masking factor [9]

$$S_h(m,n) = e^{\{-0.04 V_h(m,n)\}} \tag{21}$$

is defined in terms of a horizontal local activity function

$$V_h(m,n) = \frac{|i(m,n-1) - i(m,n+1)|}{2}. \tag{22}$$

Defining the vertical masking factor $S_v(m,n)$ similarly, we compute the masked error at each pixel as

$$f_5(m,n) = I_M(m,n)|e_w(m,n)|(S_h(m,n) + S_v(m,n)), \tag{23}$$

where $I_M(m,n)$ is an indicator function which selects pixels *close* to high intensity transitions. Note that the masking factors can be substantially less than 1 in highly active regions of the image.

The final factor is now computed as

$$F_5 = \frac{1}{N_K} \sum_{m,n} f_5(m,n) \tag{24}$$

where N_K is the number of pixels whose 3×3 Kirsch edge response $k(m,n) \geq K$, for a threshold $K = 400$. The indicator function $I_M(m,n)$ of (23) selects the set of all pixels within $l = 4$ pixels of those pixels detected by the Kirsch operator. F_5 thus measures, with a suitable weight to account for visual masking [9], these large localized errors.

2.5 Principal Component Analysis

The distortion factors were defined so as to quantify specific types of impairments. Clearly, some of the local image impairments will contribute to several or all factors, and the factors $\{F_1, \ldots, F_5\}$ will be correlated.

A principal component analysis is carried out to quantify this correlation between distortion factors. We compute the covariance matrix

$$C_F = E\{(\bar{F} - \mu_F)(\bar{F} - \mu_F)^T\} \tag{25}$$

where $\bar{F}$ is the vector of distortion factors and μ_F is its mean.

The matrix of eigenvectors will diagonalize the covariance matrix C_F. The eigenvalues $\{\lambda_j\}$ also indicates the relative contributions of the transformed vectors, or principal components, to the total energy of the vector $\bar{F}$. The "eigenfactors" are now uncorrelated and, as we shall show, are more effective and robust in the objective assessment of image quality.

2.6 Computation of PQS

We compute the PQS quality metric as a linear combination of principal components Z_j as

$$PQS = b_0 + \sum_{j=1}^{J} b_j Z_j \tag{26}$$

where the b_j are the partial regression coefficients which are computed using Multiple Regression Analysis (MRA) [19] to fit (26) to the mean opinion scores of observers that are obtained experimentally in quality assessment tests.

In order to illustrate visually the five distortion factor images, they have been evaluated for the image "Lena" and for a JPEG coder with a (low) quality parameter setting of 15. To illustrate the relative spatial contributions of $f_i(m,n)$ to F_i, and to differentiate the factors from one another, we show in Figure 2 the original image and each of the $f_i(m,n)$ suitably magnified. We observe that, as compared to $f_1(m,n)$, $f_2(m,n)$ discards a number of small errors which occur in the flat portions of the image. $f_3(m,n)$, restricted to block boundaries, is quite high in active portions of the image and also near high contrast intensity transitions, where it will be the most visible. The structured error $f_4(m,n)$ compared, to $f_5(m,n)$ shows that structured errors are very common, and that they do not consistently coincide with image regions where visual masking occurs.

3 Visual Assessment Tests

We now turn to the experimental determination of the subjective mean opinion score for each of the encoded images.

3.1 Methods

The visual or subjective evaluation of image quality has drawn attention of a number of researchers for many years, principally in relation to the evaluation of new transmission or coding schemes, and in the development of advanced television standards.

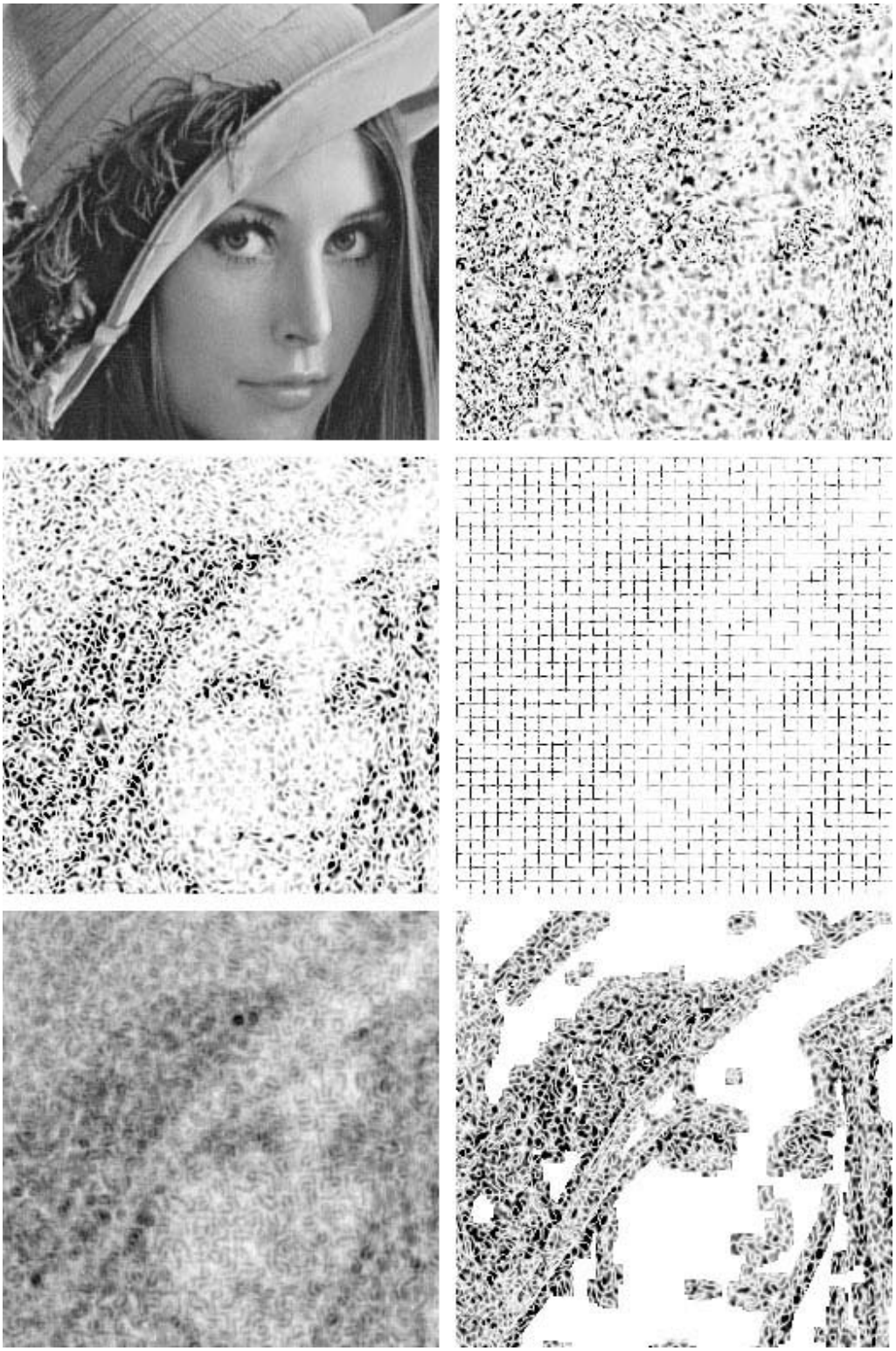

Figure 2: PQS factor images. Left to right, then top to bottom: original, $f_1(m,n)$, $f_2(m,n)$, $\sqrt{f_{3h}^2(m,n) + f_{3v}^2(m,n)}$, $f_4(m,n)$, and $f_5(m,n)$.

Scale	Impairment
5	Imperceptible
4	Perceptible, but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

Table 1: MOS grading scale.

As it applies to television, an excellent presentation of the complex issues involved has been given by Allnat [20]. The standardization committees of the ISO, and in particular the CCIR, has published recommendations on the assessment of picture quality in television. In our work, we follow closely the CCIR 500 recommendations with respect to subjective scales and experimental conditions [2].

In Table 1, we show the 5 point (MOS) impairment scale and in Table 2, the conditions for subjective assessment tests recommended in CCIR 500. The specific conditions used were as follows:

1. The pictures used were all 256×256 pixels and were viewed at 4 times the picture height ($4H$).

2. The selected observers, principally graduate students at JAIST, the Japan Advanced Institute of Science and Technology, received limited training. The selection of subjects was based on consistency in the evaluation of picture quality. The training consisted of the description and illustration of the types of distortions that the subjects would observe.

3. The subjects were instructed to grade the image quality in 1/2 step increments.

A number of coded images were evaluated informally at various times during the study, with more than 800 coded image evaluations performed, in the preliminary and final assessment tests. The results reported are based on the 675 quality evaluations of 75 encoded images by the pool of nine subjects.

3.2 Test Pictures

The image impairments represented by the distortion factors F_1 and F_2 are mainly global distortions, while F_3, F_4, and F_5 characterize local distortions. Local distortions are apparent only in portions of the images and depend on the density of transitions and flat regions in the images. Five test images were used that represent a range of characteristics. These images, shown in Figure 3, include the ITE (Institute of Television Engineers of Japan) test images "Church", "Hairband" and "Weather" [21], and the widely used "Barbara", and "Cameraman" images.

Ratio of viewing distance to picture height	4
Room illumination	None
Peak luminance on the screen	42.5 (cd/m^2)
Lowest luminance on the screen	0.23 (cd/m^2)
Time of observation	unlimited
Number of observers	9 (expert observers)

Table 2: Conditions of the subjective assessment tests.

Figure 3: PQS test images. From left to right, then top to bottom: Church, Hairband, Weather, Barbara, and Cameraman.

3.3 Coders

There are a large variety of coders, such as DPCM, Orthogonal Transform Coders (OTC), VQ, subband coders, etc. We have concentrated on DCT based JPEG, as well as widely used wavelet and subband coding technqiues. The subband coder used in our experiments is a 28 band decomposition using Johnston's 16 tap filters [22] and the wavelet coder, a 10 band decomposition with 8 tap Daubechies' filters [].

	F_1	F_2	F_3	F_4	F_5
F_1	1.00000	0.99671	0.97530	0.83914	0.58231
F_2	0.99671	1.00000	0.97432	0.82308	0.56825
F_3	0.97530	0.97432	1.00000	0.87491	0.56958
F_4	0.83914	0.82308	0.87491	1.00000	0.67142
F_5	0.58231	0.56825	0.56958	0.67142	1.00000

Table 3: Covariance Matrix

3.4 Determination of MOS

The observers are asked to assign a score $A(i,k)$ to each encoded image, where $A(i,k)$ is the score given by the i^{th} observer to image k. Each score is chosen from $\{1.0, 1.5, \ldots, 5.0\}$ according to the impairment scale of Table 1. For each encoded image, the scores are averaged to obtain the MOS value for a specific image,

$$MOS(k) = \frac{1}{n}\sum_{i=1}^{n} A(i,k) \tag{27}$$

where n denotes the number of observers.

4 Results of Experiments

The experiments resulted in a set of five images coded with one of three types of coders and for the entire range of quality. A total of seventy five encoded images were assessed by nine observers as described, and the average MOS score was computed for each encoded image. The set of error images were then analyzed.

4.1 Principal Component Analysis

The set of error images was first used to compute the distortion factors. From this set of distortion factors, we compute the covariance matrix C_F of Table 3. A principal component analysis of C_F is then carried out. We show in Table 4, the eigenvectors and eigenvalues of matrix C_F .

Note the very high correlation between F_1 and F_2. This is expected, since these factors both evaluate random errors, with some changes in their weighting. The high correlation of F_3 with F_1 and F_2 is more surprising. Although the spatial contributions to these factors are distinct, the high correlation indicates that, when aggregated into a single number, they do track each other as the coding parameters change.

The resulting eigenvalues have a wide spread of values, and the largest 3 eigenvalues amount for 99.5% of the total energy. The space spanned by the 5 distortion factors is essentially three dimensional. The eigenvectors ℓ_1, ℓ_2, ℓ_3 provide a useful

λ_1	λ_2	λ_3	λ_4	λ_5
4.191650	0.591437	0.190214	0.023917	0.002781

ℓ_1	ℓ_2	ℓ_3	ℓ_4	ℓ_5
0.475001	-0.244488	0.270834	-0.387094	0.701004
0.471985	-0.268292	0.323841	-0.310442	-0.709931
0.475262	-0.247235	0.019208	0.842713	0.049659
0.450578	0.105979	-0.862187	-0.200654	-0.046043
0.350300	0.892126	0.279357	0.057934	-0.002157

Table 4: Eigenvalues and corresponding eigenvectors.

first transformation of the F_i into an effective principal component representation (Z_1, Z_2, Z_3). To obtain a numerical distortion value, we carry out a multiple regression analysis between the principal component vector and the measured MOS values.

4.2 Multiple Regression Analysis

The partial regression coefficients b_0 and $\{b_j\}$ for the three principal components, $Z_1 - Z_3$, have been evaluated, so that PQS for any coded image is given by

$$PQS = 5.632 - 0.068Z_1 - 1.536Z_2 - 0.0704Z_3 \tag{28}$$

Note that the PQS, that is derived using principal components, can be also be expressed in terms of the distortion factors, F_i.

$$PQS = 5.797 + 0.035F_1 + 0.044F_2 + 0.01F_3 - 0.132F_4 - 0.135F_5 \tag{29}$$

It is important to use the principal components in the multiple regression analysis and the determination of PQS. Since the covariance matrix is nearly singular, the results obtained by multiple regression with the F_i or with the entire set Z_i is unstable. Although a slightly better fit of the specific test data is then obtained, the regression coefficients are not robust, and the results not usable for images outside the test set.

4.3 Evaluation of PQS

A fairly good agreement between the PQS and MOS is achieved, as shown in the scatter diagram of Figure 4. Note that the goodness of fit is better in the middle of the quality range than at its extremes. This observation is elaborated in section 6. In order to describe the degree of approximation of PQS to MOS quantitatively, the correlation coefficient R between PQS and MOS is evaluated [19]. The correlation coefficient $R = 0.928$, which is a great improvement when compared to the correlation of $R = 0.57$ of the conventional WMSE scale which is calculated using F_1 alone.

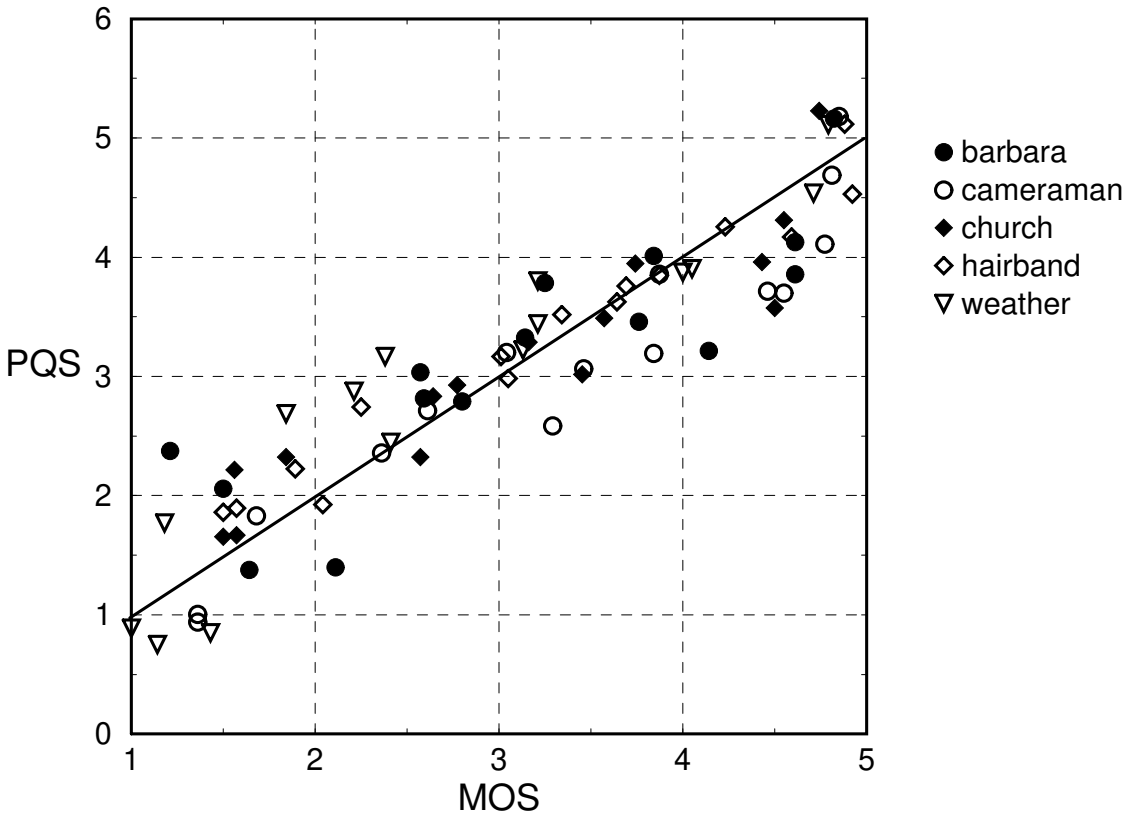

Figure 4: PQS versus MOS scatter diagram.

4.4 Generality and Robustness of PQS

We have already discussed the importance of using principal component analysis in the determination of PQS. Other issues related to the generality and robustness of PQS is its use for image data not in the test set, which is critical to any application. To assess this feature, the complete PQS evaluation was computed on each set of 4 images, and the resulting formulas used for the fifth image. We find that the results with respect to regression coefficients and correlation with MOS are very close to the values reported. We also evaluated the effect of the encoding technique on the results. We find that the PQS evaluation is only slightly dependent on the encoding technique.

5 Key Distortion Factors

We have defined the distortion factors F_i as measures of perceived disturbances which are common and basic to coding techniques. We now consider the interpretation of the combinations of F_i's into principal components, and whether each F_i, as defined, is a key distortion factor so as to further remove redundancies or to rank the importance of the factors in the set $\{F_i\}$.

5.1 Characteristics of the Principal Components

The eigenvectors $\{\ell_k\}$ that were used to compute $\{Z_j\}$ are indicated earlier. We also found that the characterization of the overall distortion is concentrated into three principal components. Let us consider the characteristics of ℓ_1, ℓ_2 and ℓ_3.

- Z_1 may be reasonable named "the amount of error," since the entries in the first eigenvector ℓ_{1i} in Table 4 are almost equal to each other.
- Z_2 may be named "the location of error," because ℓ_{52} which measures the contribution of errors in the vicinity of image edges or transitions is very large (ℓ_{52}=0.89).
- Z_3 may be named "the structure of error," because of the weight given to the factor F_4 in the third eigenvector.

As a conclusion, it is considered that PQS is given by the linear combination of three important essential factors of distortion: "the amount of error", "the location of error" and "the structure of error".

5.2 Contribution of the Distortion Factors

We now evaluate the importance of the distortion factors, taken singly and in combination. For all sets of combinations of F_i, the correlation coefficients R and R^* are shown in Table 5, where R^* is the modified R [19] adjusted for the degrees of freedom. Specifically,

$$R^* = \sqrt{\frac{R^2(n-1)-p}{n-p-1}} \tag{30}$$

where n is the total number of coded images in the test set, and p is the number of factors F_i retained. Note that, as done in the previous section, a principal component analysis was performed on the covariance matrix of the retained coefficients. The largest eigenvalues and corresponding eigenvectors were used until their accumulative value exceeded 99%. We find that the four factors, F_1, F_2, F_3 and F_4 only span a 2 dimensional space because of the high correlation between the first three factors.

Factors					R	R^*	$E\{\|error\|\}$
1					0.5665	0.5582	0.8347
	2				0.5373	0.5282	0.8510
		3			0.5714	0.5633	0.8269
			4		0.7426	0.7384	0.6551
				5	0.8985	0.8971*	0.4268
1	2				0.5547	0.5461	0.8418
1		3			0.5729	0.5564	0.8289
1			4		0.7499	0.7417	0.6362
1				5	0.9001	0.8972	0.4206
	2	3			0.5779	0.5616	0.8222
	2		4		0.7539	0.7459	0.6336
	2			5	0.8991	0.8962	0.4236
		3	4		0.7600	0.7522	0.6302
		3		5	0.9015	0.8986	0.4144
			4	5	0.9180	0.9156*	0.3773
1	2	3			0.5716	0.5550	0.8265
1	2		4		0.7516	0.7435	0.6352
1	2			5	0.8996	0.8967	0.4220
1		3	4		0.7600	0.7522	0.6302
1		3		5	0.9015	0.8986	0.4144
1			4	5	0.9271	0.9239	0.3536
	2	3	4		0.7600	0.7522	0.6302
	2	3		5	0.9015	0.8986	0.4144
	2		4	5	0.9302	0.9271*	0.3460
		3	4	5	0.9279	0.9247	0.3551
1	2	3	4		0.7600	0.7522	0.6302
1	2	3		5	0.9015	0.8986	0.4144
1	2		4	5	0.9285	0.9253*	0.3501
1		3	4	5	0.9279	0.9247	0.3551
	2	3	4	5	0.9279	0.9247	0.3551
1	2	3	4	5	0.9279	0.9247	0.3551

Table 5: Correlation Coefficient Between PQS and MOS

We also show on Table 5 the average absolute error difference between the PQS and the MOS scores assigned to the encoded images. Table 5 can be examined in several ways. First, we rank the importance of the factors taken in groups. The most important single factor is F_5. For two factors, we would use F_4 and F_5; for three F_2, F_4 and F_5: for four $F1$, F_2, F_4 and F_5.

The importance of the factor F_5 by itself stands out in the examination of Table 5, and suggests that it could be used as a quality metric by itself. Note however that a study of the relation between MOS and PQS, when F_5 alone is used as a metric, indicates that such a metric could be used at high quality, but that its performance is much worse than the PQS, given in (29), at lower quality.

Given that the subjective assessment of test images has only a precision of 0.5, an average absolute error of less than 0.35 seems adequate, but we note that the maximum error observed is as high as 0.94 for the best choice, and occurs for the extreme low end of the quality range. In particular, for F_5 used by itself as a metric, this maximum error can be as high as 1.4.

We found also that the relative contribution of the first three distortion factors to the PQS value decrease for increasing quality. This observation is consistent with the common experience that structured errors, principally in the vicinity of edges, are the only disturbances perceived for higher bit rates and quality, and in particular, that the end of block impairments are important only at low quality . This result also indicates that a PQS metric devised and applied only for the low end of the image quality range would lead to different weights and results, as we verified experimentally.

5.3 Other Distortion Factors

We have considered briefly other possible distortion factors which have been identified by practitioners. The purpose of such a study was to determine if the inclusion of quantitative factors for other known image impairments due to coders would improve the degree of approximation of PQS and MOS, and also, possibly, increase the applicability of PQS.

One such distortion factor would quantify the jagged distortion of smooth edges that is introduced by some image coding techniques. It is possible to define such a distortion factor by analyzing locally the discrepancies between the direction of the image edge and the direction of the edge in the coded image. We found that such a factor has a slight effect on one of the images in our test set, principally at high quality, but does not affect the overall results presented. We also observed that a moire type of disturbance occurs in the scarf portion of the test image"Barbara" at specific quality levels. This effect occurs in such isolated instances that the overall approach and results are not affected. However, this additional study pointed out some of the limitations of our results, and will be elaborated in the next section.

6 Discussion

Since this work was started some years ago, the interest and importance of visual assessment and of quality metrics for image processing and image coding has increased greatly. In this section, we consider first some of the limitations and promise of the methodology presented here for applications of recent interest. Second, we discuss

the components of an overall strategy for systematic advances in perceptually based image coding.

6.1 Limitations in Applications

The set of distortion factors that we have defined span a three dimensional space. Thus, image impairments are characterized by a three dimensional vector. Multiple regression analysis, a statistical regression technique, has allowed the reduction of this vector to a single number, PQS, which has a good correlation with MOS. In this statistical fit, distortion factor contributions may be positive or negative. If these factor contributions are outside the range for which the regression was carried out, the resulting PQS may be invalid. Thus, for very poor quality images, it is possible to obtain negative values of PQS, a meaningless result. For the PQS value to be meaningful, one requirement is that the weighted contributions of each of the factors as given in (29) be in the range 1 to 5. From the scatter diagram of Figure 4, we observe that PQS provides the best match to MOS values in the middle of the quality range. We could improve the performance of PQS by using a piecewise linear model, or by two separate measures, one for low to medium quality images and one for medium to high quality . We have performed a complete analysis for the lower quality range, and find that the contributions of the distortion factors are then substantially different than what we reported for a fit over the entire quality range. The predictive value of this limited range PQS is also improved. These observations are related to the choice of subjective quality assessment scales and performance ranges discussed next.

6.2 Visual Assessment Scales and Methods

The visual assessment methods of CCIR 500 were developed for use in entertainment broadcast television, i.e. of fairly high quality, and are targeted to non expert viewers for images and video with low information content. Image coding applications now have a much wider range. The subjective 5 point impairment scale for images or video of CCIR 500 is too broad and not precise enough for many current applications. For instance, for the video conferencing of head and shoulder images, acceptable image quality is much lower than for broadcast television and still image coding. Thus all such encoded images and video will cluster at the low end of the impairment scale. The same comment applies to high quality image coding, where the preservation of critical details is important. Thus, alternate quality scales, and a different subjective assessment method, such as using anchor images, are needed [20, 23]. The introduction of adaptive coding techniques, where the bit rate constraints dynamically modify some of the quantization parameters, such as in MPEG2, has also led to the use of a variable and continuous rating of quality [3].

6.3 Human Vision Models and Image Quality Metrics

In our work, we have made use of a simple global or single channel model of visual perception. Much progress has been made in the development of multichannel models of visual perception [24, 25, 26] and in the careful determination of the visual masking thresholds for each of the channels [27]. Such threshold models have been used by Daly [28] in the determination of a Visual Distortion Predictor (VDP) that computes the probability of detection of a visible error. These perceptual models have application to the evaluation of quality in the processing of very high quality images, where errors are small and close to threshold. We have compared PQS to the VDP for high quality image coding. The distortions images predicted by both methods were quite similar for high quality [29]. Additional quantitative work for high quality image coding, that combines more complete perceptual models and the PQS methodology, is now under way at CIPIC. The promise of metrics based strictly on perceptual models is that they would apply to all types of image impairments, and not only to image coding. The limitation of such an approach, besides its restricted applicability to impairments at the visual threshold, is in understanding or controlling image display and viewing conditions to fit the models. Note that successful limited use of properties of visual perception for image processing or encoding applications have been reported through the years [30, 31, 32, 33, 34, 35].

6.4 Specializing PQS for a Specific Coding Method

We have proposed a methodology for a "general" PQS which will give useful results for different types of coders and images. When the coding method is fixed and the test image data and the coder are fixed, we can tune the picture quality scale to that coder and obtain then more accurate results [36]. Such a specialization is useful for coder design or adaptive coding parameter adjustment [14].

6.5 PQS in Color Picture Coding

The PQS methodology has been applied to color picture coding. For color, the importance of a perceptually uniform color space leads us to consider color differences as defined in the Munsell Renotation System which is psychometrically uniform and metric. Then, we can utilize the color differences instead of $e(m, n)$ for a PQS applicable to coded color images [37, 38].

7 Applications of PQS

There a number of applications for a perceptually relevant distortion metric such as PQS. Coding techniques can now more confidently be compared by checking their PQS values for the same bit rate, or bit rates for the same image quality. For instance,

for wavelet coders, alternatives in the choice of wavelets, quantization strategies, and error free coding schemes have been compared [15, 16]. This comparison make possible a systematic choice of parameters, on the basis of a meaningful measure. The same study shows that SNR does not help in making such choices.

Optimization of quantization parameters in coders based on PQS has also been examined [14]. For adaptive coders, some of the factors, $\{F_i\}$, can be estimated locally by a decoder within the coder. Hence, the objective picture quality metric PQS that measures the degradation of quality in the coded picture can then be reduced by adjustments of the coder parameters [14].

8 Conclusions

We have proposed a methodology for devising a quality metric for image coding and applied it to the development of an objective Picture Quality Scale (PQS) for achromatic still images. This PQS metric was developed over the entire range of image quality defined by the impairment scale of CCIR 500. PQS is defined by taking into account known image impairments due to coding, and by weighting their quantitative perceptual importance. To do so, we use some of the properties of visual perception relevant to global image impairments, such as random errors, and emphasize the perceptual importance of structured and localized errors. The resulting PQS closely approximates the Mean Opinion Score (MOS), except at the low end of the image quality range. We have also interpreted the PQS system as composed of a linear combination of three essential factors of distortion: the "amount of error", the "location of error' and the "structure of error". We have also discussed some of the extensions and applications of such an objective picture quality metric.

Systematic studies of the objective evaluation of images, as well as their subjective assessment, are difficult and only now becoming active areas of research [4]. However, the rapid increase in the range and use of electronic imaging and coding and their increasing economic importance justifies renewed attention and specialization of perceptually relevant image quality metrics as a critical missing component for systematic design and for providing the quality of service needed in professional applications.

References

[1] N. W. Lewis and J. A. Allnatt, "Subjective quality of television pictures with multiple impairments," *Electronic Letters*, vol. 1, pp. 187–188, July 1965.

[2] CCIR, "Rec. 500-2. Method for the subjective assessment of the quality of television pictures." In, Recommendations and reports of the CCIR, Geneva, 1982.

[3] Proc. MOSAIC Workshop:, "Advanced methods for the evaluation of television picture quality," proceedings of the mosaic workshop, Institute for Perception Research, Eidhoven, Netherlands, 1995.

[4] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, vol. 81, pp. 1385–1422, Oct. 1993.

[5] CCIR, "Rec. 567-1. Transmission performance of television circuits designed for use in international connections, pl-38." In Recommendations and reports of the CCIR and ITU, Geneva, 1982.

[6] T. Fujio, "Two-dimensional processing of TV signal," Tech. Rep. Vol. 25, NHK Technical Research Laboratory, Mar. 1973.

[7] R. A. Bednarek, "On evaluation impaired television pictures by subjective measurements," *IEEE Transactions on Broadcasting*, vol. BC-25, no. 2, pp. 41–46, 1979.

[8] F. X. J. Lukas and Z. L. Budrikis, "Picture quality prediction based on a visual model," *IEEE Transactions on Communications*, vol. COM-30, pp. 1679–1692, July 1982.

[9] J. O. Limb, "Distortion criteria of the human viewer," *IEEE Transactions on System, Man, and Cybernetics*, vol. SMC-9, pp. 778–793, Dec. 1979.

[10] D. J. Sakrison, "On the role of the observer and a distortion measure in image transmission," *IEEE Transactions on Communications*, vol. COM-25, Nov. 1977.

[11] A. Oosterlick *et al.*, *Image Coding Using the Human Visual System*, ch. 5. Katholieke Universiteit: Katholieke Universiteit, 1987.

[12] C. H. Graham *et al.*, *Vision and Visual Perception*. John Wiley and Sons, Inc., 1965.

[13] M. Kunt *et al.*, "Second-generation image-coding techniques," *Proceedings of the IEEE*, vol. 73, pp. 549–574, apr 1985.

[14] V. R. Algazi, G. E. Ford, M. Mow, and A. Najmi, "Design of subband coders for high quality based on perceptual criteria," in *Proceedings of SPIE, Application of Digital Image Processing XVI*, vol. 2028, (San Diego, California), pp. 40–49, SPIE, 1993.

[15] J. Lu, V. R. Algazi, and R. R. Estes, Jr., "A comparative study of wavelet image coders," *To appear in Optical Engineering*, Sept. 1996.

[16] J. Lu, V. R. Algazi, and R. R. Estes, Jr., "A comparison of wavelet image coders using a picture quality scale," in *Proceedings of SPIE Wavelet Applications Conference*, (Orlando, Florida), SPIE, Apr. 1995.

[17] Y. Horita and M. Miyahara, "Image coding and quality estimation in uniform perceptual space," IECE Technical Report IE87-115, IECE, Jan. 1987.

[18] 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, pp. 713–724, June 1995.

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

[20] J. Allnatt, *Transmitted-picture Assessment*. John Wiley and Sons, Inc., 1983.

[21] ITE, "ITE television system chart (version II)," tech. rep., Digital Standard Picture, ITE, Dec. 1987.

[22] J. D. Johnston, "A filter family designed for use in quadrature mirror filters," in *Proceedings of IEEE ICASSP 1980*, pp. 291–294, IEEE, 1980.

[23] J. B. M. A. M. van Dijk and A. B. Watson, "Quality assessment of coded images using numerical category scaling," in *Advanced Image and Video Communications and Storage Technologies, Amsterdam*, vol. 2451 of *SPIE*, pp. 90–101, 1995.

[24] S. J. P. Westen, R. L. Lagendijk, and J. Biemond, "Perceptual image quality based on a multiple channel HVS model," in *International Conference on Acoustics, Speech, and Signal Processing*, vol. 4 of *ICASSP 95*, pp. 2351–2354, ICASSP, 1995.

[25] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *International Conference on Image Processing*, vol. 2, pp. 982–986, ICIP, 1994.

[26] C. Zetzsche and G. Hauske, "Mustiple channel model for the prediction of subjective image quality," in *Society of Photo Instrumentation Engineering Proceedings*, vol. 1077, pp. 209–216, SPIE, 1989.

[27] E. A. B. Watson, *Digital Image and Human Vision*. MIT Press, 1993.

[28] S. Daly, *The Visible Difference Predictor: An Algorithm for the Assessment of Image Quality, in Digital Image and Human Vision*, ch. 14. MIT Press, 1993.

[29] N. Avadhanam and V. R. Algazi, "Prediction and measurement of high quality in still image coding," in *Proceedings of SPIE Very High Resolution and Quality Imaging Conference, EI 96*, vol. 2663, (San Jose, California), pp. 100–109, 1996.

[30] V. R. Algazi, G. E. Ford, and H. Chen, "Linear filtering of images based on properties of vision," *IEEE Transactions on Image Processing*, vol. 4, pp. 1460–1464, Oct. 1995.

[31] T. A. Hentea and V. R. Algazi, "Perceptual models and the filtering of high-contrast achromatic images," *IEEE Transactions Systems, Man and Cybernetics*, vol. SMC-14, pp. 230–246, Mar. 1984.

[32] A. N. Netravali, "Interpolative picture coding using a subjective criterion," *IEEE Transactions on Communications*, vol. COM-25, pp. 503–508, May 1977.

[33] J. O. Limb, "On the design of quantizer for dpcm coder: A functional relationship between visibility, probability, and masking," *IEEE Transactions on Communications*, vol. COM-26, pp. 573–578, May 1978.

[34] B. Girod *et al.*, "A subjective evaluation of noise-shaping quantization for adaptive intra-/interframe dpcm coding of color television signals," *IEEE Transactions on Communications*, vol. COM-36, pp. 332–346, Mar. 1988.

[35] R. Wilson *et al.*, "Anisotropic non-stationary image estimation and its applications: Part ii - Predictive image coding," *IEEE Transactions on Communications*, vol. COM-31, pp. 398–406, Mar. 1983.

[36] M. Miyahara, "Quality assessments for visual service," *IEEE Communications Magazine*, vol. 26, pp. 51–60, Oct. 1988.

[37] M. Miyahara and Y. Yoshida, "Mathematical transformation of (R,G,B) color data to munsell (H,V,C) color data," in *SPIE Visual Communications and Image Processing III*, pp. 1001–1118, Nov. 1988.

[38] K. Kotani, Q. Gan, M. Miyahara, and V. R. Algazi, "Objective picture quality scale for color image coding," in *Proceedings of International Conference Image Processing, ICIP 95*, vol. 3, (Washington D. C.), pp. 133–136, Oct. 1995.