Meeting of experts on subjective and objective video quality

14-16 October 1997

Objective Picture Quality Scale for Digital Compressed Picture for Broadcast*[1]

NHK Science & Technical Research Laboratories

1. Introduction

Considering the trend towards adopting high-efficiency picture coding schemes into digital broadcasting services, an objective picture quality scale for evaluating digital compressed pictures has been investigated. This scale is derived from the following three distortion factors; 1) weighted noise by the spatial and temporal frequency characteristics of human vision, 2) weighted noise by the masking effects of human vision, and 3) block distortion. Furthermore, the degree of these distortions varies over time, yet subjective evaluations give a single score for picture quality that varies over the particular presentation time. Therefore, methods for determining a single objective picture quality value for time-varying distortion is also investigated. As a result, a generally applicable objective picture quality scale is obtained that correlates well with subjective picture quality scale for standard TV pictures coded by hybrid DCT scheme, irrespective of the contents of the pictures.

2. Definitions of three types of distortion

2.1 Three-dimensional(3D) weighted noise

The three-dimensional weighted noise $g_3(x, y, t)$ is defined in Eq. (1), which is weighted by both the spatial and temporal vision frequency characteristics.

$$g_3(x, y, t) = FFT^{1}\,[FFT\,[ph(x, y, t) - po(x, y, t)]\; V_1(Wx, Wy)\; V_2(Wt)] \qquad (1)$$

x, y, t: positions in the horizontal, vertical, and time directions
Wx, Wy, Wt: frequencies in the horizontal, vertical, and time directions
$po(x, y, t)$: original picture data
$ph(x, y, t)$: evaluated picture data
$FFT[\;]$: Three dimensional Fast Fourier Transform
$FFT^{1}[\;]$: Three dimensional Inverse Fast Fourier Transform
$V_1(Wx, Wy)$: Spatial human vision frequency characteristics
$V_2(Wt)$: Temporal human vision frequency characteristics

Here, Eq. (2) from Ref.[1] is used for the spatial vision frequency characteristics, reported to be the most valid for standard TV pictures. For the temporal characteristics Eq. (3) from Ref.[2].is used.

---

*[1] This document has been presented at the last WP1/9 meeting as a part of a contribution from Japan, and is a summary of the following paper:

Hamada, H. and Namba, S., "A Study on Objective Picture Quality Scales for Pictures Digitally Encoded for Broadcast", IEICE Transactions on Communications, Vol.E77-B, No.12, December 1994.

$$V_1(Wx, Wy) = 2.46[0.1+0.25f]\exp[-0.25f] \quad (2)$$
$$f = (Wx^2 + Wy^2)^{1/2}$$

$$V_2(Wt) = 0.134[1+(Wt / 0.5)]/[1+(Wt / 7.8)^2]^{1.2} \quad (3)$$

When computing the FFT for five seconds of picture data (720 pixels × 240 lines × 300 fields), the central portion 540 pixels × 180 lines is divided into 60-field segments and processed sequentially to shorten computer calculation time. By taking the mean-square (in units of fields) of the 3D weighted noise $g_3(x, y, t)$ obtained from this method, the distortion factor $F1(t)$, as shown in Eq. (4), is obtained using the time (in units of fields) as a parameter.

$$F1(t)=\Sigma\Sigma g_3(x, y, t)^2/N \quad (4)$$

N: The number of pixels in a field

An example of the 3-D WSNR over time is shown in Figure 1.

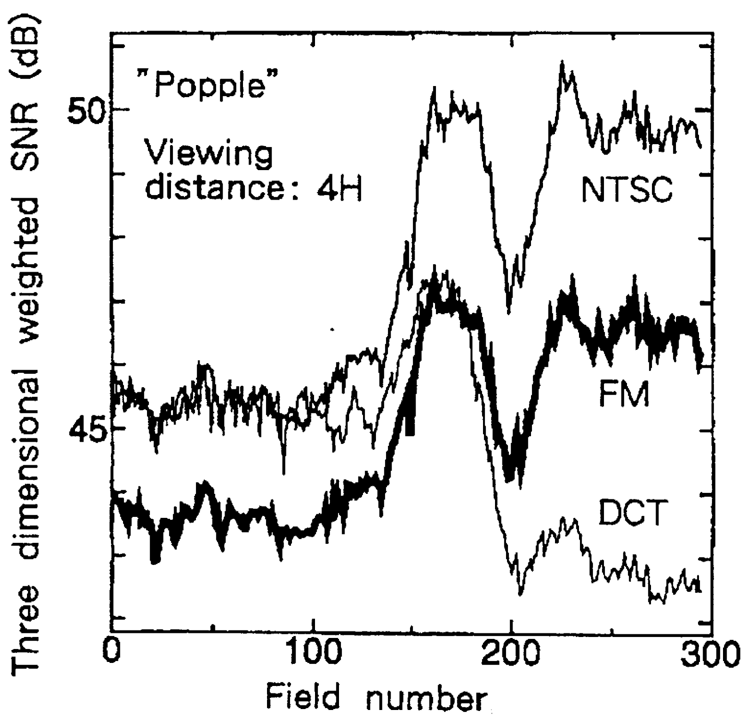

Figure 1. SNR weighted by human visual spatial-temporal frequency characteristics for "Popple."

2.2 Three-dimensional masking weighted noise

Also examined is the masking effect, where the visibility of noise varies with the degree of local activity in the picture. Here, the masking property described in Ref.[3] is used which introduced the masking effect in order to make the most appropriate quantization table for the coefficients of orthogonal transformations such as DCT. That is, the degree of local activity S within the 8-pel × 8-line block is defined as in Eq. (5). The degree to which the noise is visually noticeable is weighted by $F(S)$. The values of S and the weighting $F(S)^{1/2}$

are shown in Table 1.

$$S = 1/64\Sigma\Sigma(u_{ij} - u_m)^2 \quad (5)$$

u_{ij}: pixel value in original picture

u_m: average value of u_{ij} within the block

Table 1. Relation between block activity and weighting value

S: Block activity	$F(S)$: Weighting value
0 - 25	0.6303
25 - 548	0.2107
548 - 1767	0.1622
1767 - ∞	0.1422

Here, first, the local activity S is obtained from the 7-pel × 7-line block of original pictures centered on each pixel. Next, the 3D weighted noise in each pixel is weighted by $F(S)^{1/2}$ to obtain the distortion $gm_3(x, y, t)$, accounting for the masking effects. $gm_3(x, y, t)$ is given in Eq. (6). The distortion factor $F2(t)$ is given in Eq. (7).using the time (in units of fields) as a parameter.

$$gm_3\ (x, y, t) = g_3(x, y, t)\ F(S)^{1/2} \quad (6)$$

$$F2(t)= \Sigma\Sigma\ gm_3\ (x, y, t)^2/N \quad (7)$$

2.3 Block distortion in hybrid-DCT-encoded pictures

Block distortion is one of the most striking distortion factors in DCT encoded pictures. The mean-square is taken for the difference in the 3D masking weighted noise at the boundaries of the blocks over the entire picture. The distortion factor $F3(t)$ is shown in Eq. (8). Here, m and n indicates the location of the block, and $E(\ )$ indicates the mean value over the entire picture.

$$F3(t)=[\{E(\Sigma(gm_3(8m, j, t) - gm_3(8m+1, j, t))^2)\}^2 + \{E(\Sigma(gm_3\ (i, 8n, t) - gm_3\ (i, 8n+1, t))^2)\}^2]^{1/2} \quad (8)$$

3. Time variance of distortions and the objective picture quality scale

3.1 Time variance of distortions and the minimum short-time-interval average noise power

Here, a method of determining a single over-all objective quality score from the time-varying distortion factors is considered. In practice, when evaluating the quality of moving pictures whose picture quality varies over time, there are many uncertainties: do assessors evaluate based on the average picture quality over the entire time, or based on the most severe degradation? Can their eyes catch instantaneous drops in picture quality? To address this, the minimum short-time-interval average noise power is defined. This minimum short-time-interval average noise power(Δt) is defined as follows:

"The average noise power over the time interval Δt from t_0 to $t_0 + \Delta t$ is calculated for all t_0

within the total presentation time, and the lowest value is taken as the evaluation score for that picture."

An example of the minimum short-time-interval average noise power obtained from the 3D weighted noise power is shown in Figure 2. It is converted and expressed as an SNR. This minimum 3D-WSNR varies according to the interval Δt. When Δt is small it approaches the instantaneous picture quality degradation, while when Δt is large, it corresponds to the average picture quality. When $\Delta t = 0$, it becomes the minimum noise power in the five-second interval, and when $\Delta t = 5$ sec., it becomes the average noise power over the entire interval.

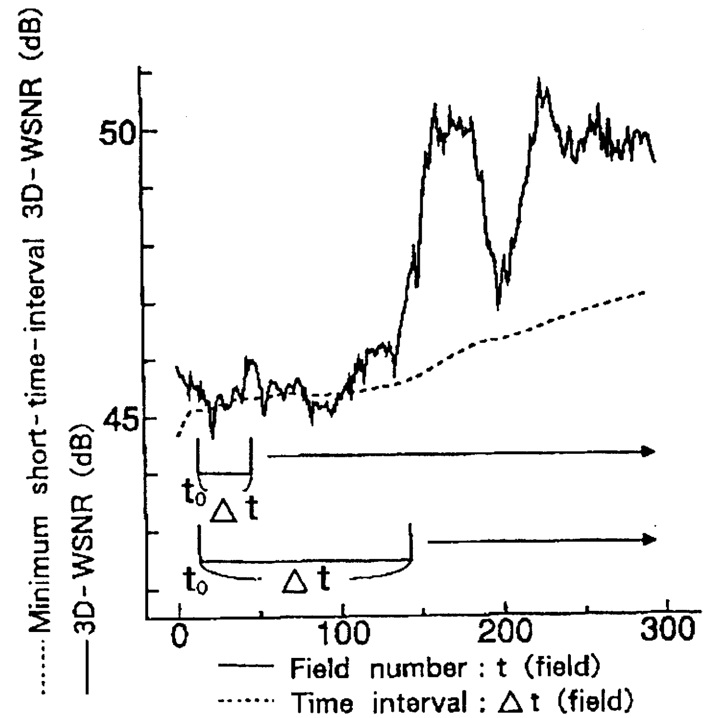

Figure 2. 3D-WSNR and minimum short-time-interval 3D-WSNR

3.2 An objective picture quality scale based on multiple regression analysis

Here, described is a method of obtaining an objective picture quality scale using the above minimum short-time-interval average noise power for each distortion factor. For three distortion factors $F1(t)$, $F2(t)$, $F3(t)$ of each picture, the minimum short-time-interval average noise power is calculated using Δt as a parameter. After converting these noise powers to the SNR, mean separation and normalizing , an objective picture quality scale is obtained through multiple regression analysis with subjective picture quality scores. That is, various linear combinations of these were taken to determine the one with the least variance from the subjective scores. For this correlation between the objective scale K and the subjective evaluation S, a multiple correlation coefficient R such as that in Eq. (9) was used.

$$R = \sigma_K / \sigma_S \qquad (9)$$

$$\sigma_K^2 = E\{(K - \mu_K)^2\},\ \sigma_S^2 = E\{(S - \mu_S)^2\}$$

$$\mu_K = E\{K\},\ \mu_S = E\{S\}$$

4. Subjective evaluation

Subjective evaluation condition is shown in Table 2. This evaluation took the five pictures, each processed in three ways, DCT, NTSC, and NTSC-FM, and evaluated each from two viewing distances, 4H and 6H, to give a total of 30 data sets. The results are shown in Figure 3.

Table 2. Subjective assessment test

Evaluation condition	Conform to ITU-R Rec. BT.500
Viewing distance:	4H and 6H
Evaluation method	Double-Stimulus Continuous-Quality Scale method
Evaluation system	Hybrid DCT encoding quality NTSC studio quality Satellite reception quality (NTSC-FM, C/N = 16 dB)
Evaluated pictures	5 pictures from the ITU-R library- "Popple", "Cheerleaders", "Flower Garden", "Mobile & Calendar", "Diva with Noise"
Presentation time	5-second scene repeated twice = 10 seconds

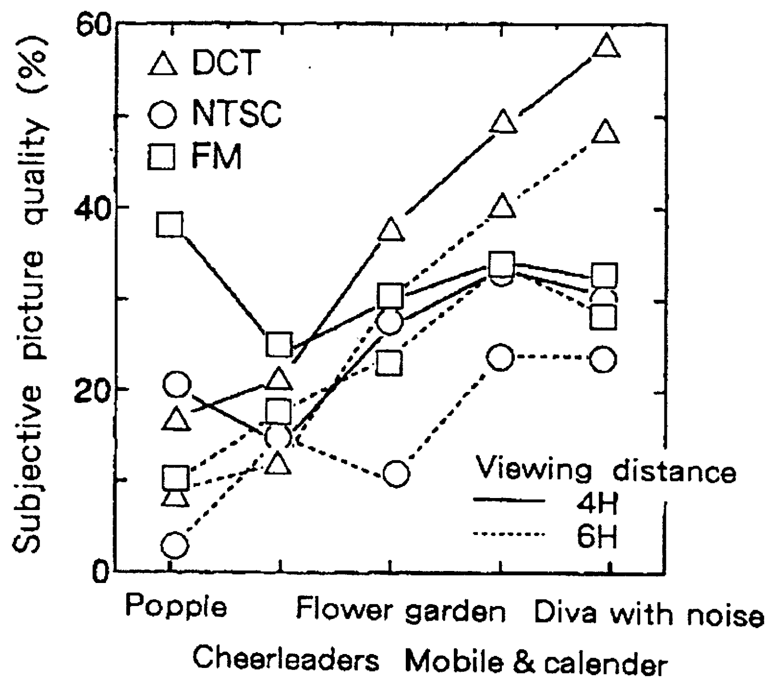

Figure 3. Subjective test results for moving pictures

5. The performance of proposed objective picture quality scale

Here, the types of distortion compared in the objective picture quality scale are:

(1) Physical noise;

(2) The 2D weighted noise, considering only the spatial frequency characteristics;

(3) The 3D weighted noise, considering both the spatial and temporal frequency

characteristics;

(4) The 3D masking weighted noise, considering the effects of masking as well;

(5) The combined 3D masking weighted noise and 3D masking weighted block distortion.

Using Δt as a parameter, an objective scale was obtained from the 30 sets of subjective evaluation data through multiple regression analysis. Its multiple correlation was calculated from the objective and subjective scales. The results obtained from distortion factors (1), (2), (3), and (4) are shown in Figure 4 as functions of Δt. As you can see, compared to the physical SNR and the 2D weighted SNR, the scale based on the 3D weighted SNR agrees much more closely with the subjective evaluations. When the masking effect is also accounted for, there is a further small but visible improvement. In addition, in all the scales, the plot of the degree of correlation against Δt is convex . This suggests that the picture should be rated not by the average SNR over the five-second interval or by the instantaneous minimum SNR, but by a value somewhere in between the two.

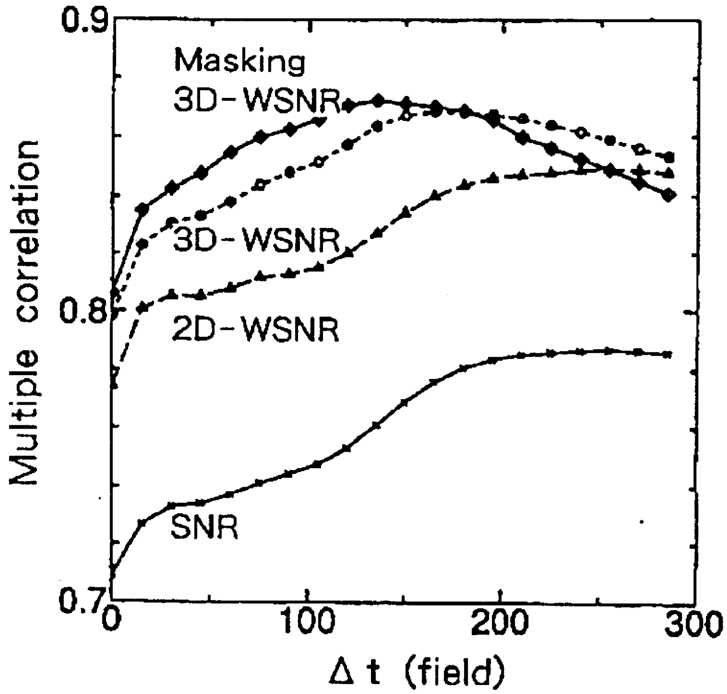

Figure 4. Comparison of objective picture quality scales derived from various distortion factors.

Next the results are investigated when block distortion was included as a distortion function. Figure 5 shows a comparison of the correlations for (4) the 3D masking weighted noise and (5) the combined 3D masking weighted noise and 3D masking weighted block distortion. Here, the data in Figure 5 is presented with the results for DCT, NTSC, and NTSC-FM pictures. Block distortion is characteristic distortion of transform encoding techniques such as DCT, so it is only natural that there is hardly any improvement in the NTSC and NTSC-FM scores. For the DCT scores, however, a definite improvement can be seen, and the objective scale considering block distortion with Δt = 165 fields (2.75 sec.) attained a high 0.95 correlation.

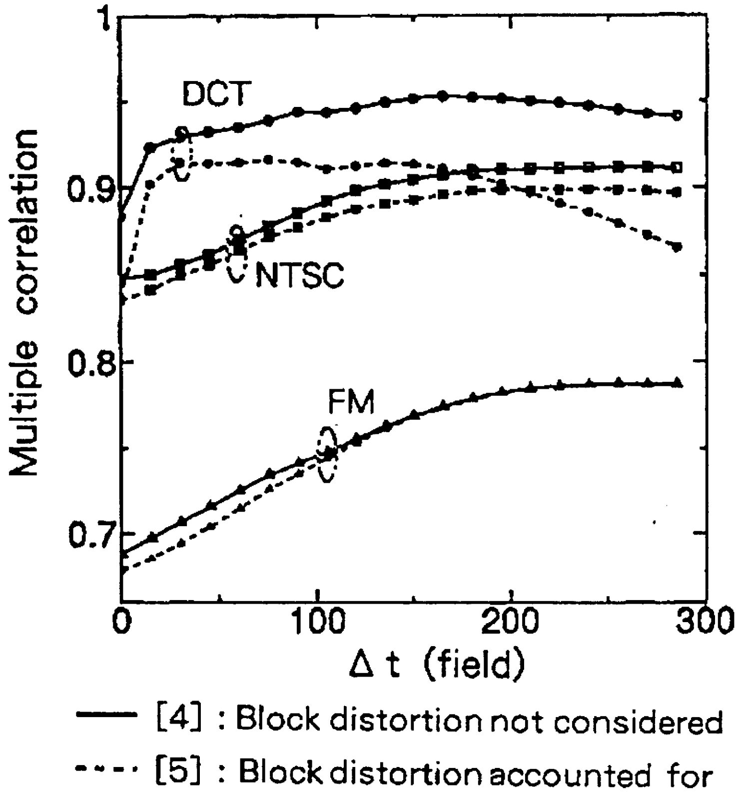

Figure 5. Effects of taking block distortion factor into account on multiple correlation for DCT, NTSC and NTSC-FM pictures.

Table 3 shows the maximum correlation values between the objective scale and subjective assessments, and gives the values of Δt for those scores. Figure 6 shows the relation between subjective evaluations and the objective scale based on (3) 3-D weighted noise for 30 pictures with Δt = 165 fields (2.75 sec.). Regardless of the pictures content, system, or viewing distance, an error is limited almost within +/- 10% (equals +/- 0.5 rank in a 5-rank scale). Figure 7 shows the relation between the objective and subjective scores for DCT pictures by means of (5), the combined 3D masking weighted noise and 3D masking weighted block distortion.

Table 3. Distortion factors and multiple correlation between subjective and objective picture quality

Distortion factors	Number of pictures	Multiple correlation	Δt (field)
1	30*1	0.802	165
2	30*1	0.853	180
3	30*1	0.868	165
4	30*1	0.872	135
4	10*2	0.916	75
5	10*2	0.953	165

*1: DCT, NTSC, NTSC-FM, *2: DCT

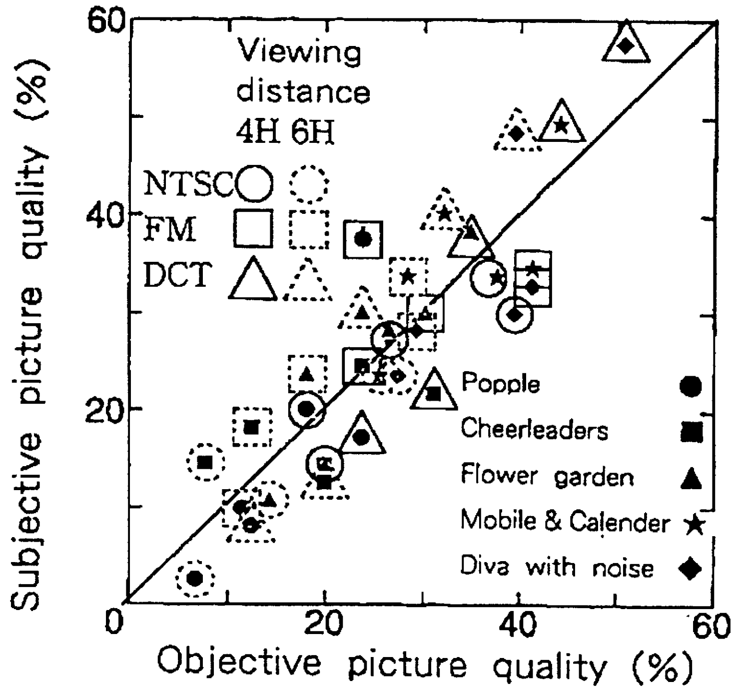

Figure 6. Relation between subjective picture quality and objective picture quality calculated by weighted SNR for DCT, NTSC and NTSC-FM moving pictures.

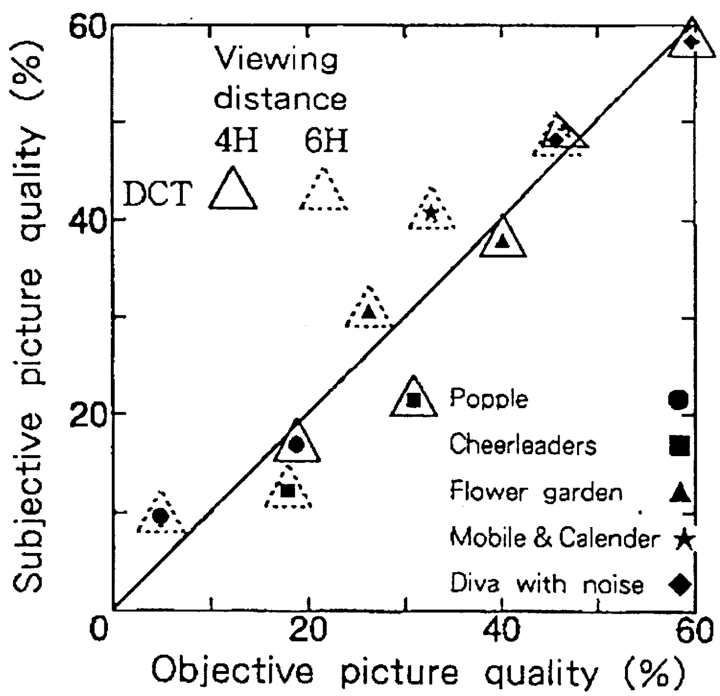

Figure 7. Relation between subjective picture quality and objective picture quality calculated by masking weighted SNR and masking weighted block distortion for DCT moving pictures.

6. Conclusion

Objective picture quality scales for digitally encoded pictures have been examined. The followings have been found:

1) Compared with an evaluation scale based on the 2D weighted SNR, further improvement was seen using a scale based on the 3D weighted SNR, considering both the temporal and spatial frequency characteristics of human vision.

2) There was improvement, albeit small, obtained by considering the masking effects.

3) Considering block distortion further improved the correlation greatly.

4) The results suggest that pictures should be rated not by the average SNR over the five-second interval nor by the instantaneous minimum SNR, but by a value in between the two.

5) For a hybrid-DCT-encoded standard moving picture, a scale combining the 3D masking weighted noise (which accounts for both temporal and spatial frequency characteristics) and block distortion achieved a 0.95 correlation.

There remain a number of issues to be considered in future studies to further refine these scales as follows:

1) Evaluation of distortion in color space

2) Evaluation of dynamic distortion such as the decrease in visual perceptiveness immediately following scene changes and mosquito noise in moving pictures

3) Evaluation of temporally and spatially localized distortion

4) High-precision matching methods for time-varying distortion and evaluation scales

5) Applicability to many different types of pictures

References

[1]Chitpraser,B., Rao,K.P., "Human visual weighted progressive image transmission", IEEE Trans., Commun., COM-38,7, pp.1040-1044, July 1990.

[2]Kubota,K., Nishizawa,T., "A three dimensional noise weighting function and its application to HDTV transmission", IEICE Trans., vol.J69-B, pp.503-511, May. 1986.

[3]Hamada,T., Matsumoto,S., "Optimal quantization scheme of orthogonal transform coefficients considering noise masking effect due to local activity", IEICE Trans., vol.J75-B-1, no.12, pp.791-801, Dec. 1992.

[4]Miyahara,M., Kotani,K., Horita,Y., Fujimoto,T., "Objective picture quality scale(PQS)-consideration of local feature and universality-", IEICE Trans., vol. J73-B-1, no.3, pp.208-218, Mar. 1973.