

A New Vision-Based Evaluation Method for Image Noise: A Computational Subjective Evaluation Method Using A Cooperative Human Vision Model

Toshikazu Matsui

*Graduate School of Engineering, Gunma University
Kiryu, Gunma, Japan*

Abstract

A new objective image noise evaluation method based on human vision's perceptual responses has been developed. The purpose of this research is to develop an objective image noise evaluation method that produces results equal to human subjective opinions. Concretely, the following two-step approach, i.e., computational subjective evaluation method, has been proposed: the first is to simulate perceptual responses to image noise by using a cooperative human vision model; the second is to design evaluation criteria based on the model's responses. The proposed method was applied to noise evaluation for halftone patch images with a uniform density outputted from digital color copying machines. The following results were obtained: (1) the vision model simulated correctly perceptual responses to the patch images' noise; (2) two kinds of evaluation criteria designed here correlated closely with the subjective evaluation score (correlation coefficient > 0.96); (3) the quality level of single-colored images was a factor to affect the performance of the copying machines used. These results suggest that the proposed method is effective in producing objective results in good agreement on subjective judgments of humans and that human vision's adaptively-changing perceptual responses are appropriately reflected in the criteria.

1. Introduction

In the field of digital hard copy where halftoning technologies are utilized, image noise has been considered as one of the most important image quality factors. Conventional noise evaluation methods such as RMS granularity, Wiener spectrum, and graininess scales¹ combining Wiener spectrum and the spatial frequency characteristic in human vision, etc., have some difficulty in producing results in good agreement on subjective judgments of humans, because human vision's adaptively-changing perceptual responses have not been sufficiently incorporated into their criteria. It is well known that the

human vision system changes its spatial frequency characteristic adaptively depending on the presented image's property. Moreover, though there also exist cases where the appearance for noise at subjective evaluation processes varies depending on viewpoints of subjects, the conventional methods cannot consider the effect of subjects' viewpoints because they are based on the Fourier transformation. In order to solve the above defects of the conventional methods and to develop objective image noise evaluation methods which agree well with subjective judgments, the following two-step approach can be a promising way: the first is to construct some framework capable of quantitatively simulating human vision's perceptual responses to image noise; the second is to design new evaluation criteria based on the framework. This can be referred to as computational subjective evaluation method.

In this paper, a new vision-based evaluation method for image noise is proposed based on the two-step approach, which is realized by a cooperative human vision model. The cooperative vision model is a mathematical model of the human vision system formulated on the basis of an idea that human vision's excellent image analysis functions are realized by a cooperation between the image processing mechanism in the brain and the image observing mechanism in the eye-optical system, and it is capable of quantitatively reproducing human vision's perceptual response characteristics that change adaptively according to the property of the presented image and the viewpoint²⁻⁴. Therefore, it is expected that the model could be a framework to simulate human vision's perceptual responses to image noise. In addition, the proposed method is applied to noise evaluation for halftone patch images with a uniform density outputted from three kinds of digital color copying machines, and its effectiveness is clarified by examining the coincidence between the objective and subjective evaluation results.

2. Cooperative Human Vision Model

Figure 1 represents a block diagram of the cooperative model consisting of four parts²⁻⁴: eye-optical system, retina part, brain part, and feedback loop for accommodation. The eye-optical system inputs a presented image based on the human vision's image observing strategy that can be described by two kinds of processes. One is that the degree of retinal image's blur, which is caused by a deviation of the focal distance of the visual lens from the just-focused condition, changes depending on the presented image's property; this process has been verified for still images.⁵⁻⁷ The other is that the size of visual field, which can be considered as being produced inevitably by the unequal distribution of visual receptor cells (cones), changes depending on the degree of the retinal image's blur: visual-field-suppressing process. Let a presented image be $f(r)$ and let the operation for producing the retinal image's blur be represented by a Gaussian low-pass filter with its point spread function $G(r, \tau_1)$ (τ_1 : blur parameter) and let the visual-field-suppressing operation be linearly approximated by a multiplication with a Gaussian window function $G(r - ra, \tau_0 - \tau_1)$ (ra : viewpoint, τ_0 : visual field parameter), then the variable signal component $gs(r)$ after the eye-optical system is expressed as follows.

$$gs(r) = fs(r) - K \cdot G(r - ra, \tau_0 - \tau_1), \quad (1)$$

$$fs(r) = G(r - ra, \tau_0 - \tau_1)fr(r), \quad (2)$$

$$K = \int_{-\infty}^{+\infty} fs(r) dr, \quad (3)$$

$$G(r, \tau) = (4\pi\tau)^{-0.5} \exp(-r^2/4\tau) \quad (4)$$

where $fr(r)$ represents the retinal image which can be calculated by convolving $f(r)$ and $G(r, \tau_1)$, and $fs(r)$ and K

represent the image suppressed by the window and the local average component, respectively.

The retina part works as a spatial band-pass filter with its point spread function $h_x(r)$ and its transfer function $H_x(w_s)$, and the output signal $g_x(r)$ expressed by the following convolution integral and the global average luminance B of the presented image are transmitted to the brain part.

$$g_x(r) = \int_{-\infty}^{+\infty} h_x(r - r') gs(r') dr', \quad (5)$$

In the brain part, the transmitted signals $g_x(r)$ and B are utilized for the perception of the shape, and at the same time, the following evaluation function I_x is calculated.

$$I_x = (1/B^2) \int_{-\infty}^{+\infty} \{1/G(r - ra, \tau_0 - \tau_1)\} \{g_x(r)\}^2 dr. \quad (6)$$

The feedback loop appropriately controls the degree of retinal image's blur so that the evaluation function takes a maximum point. Thus, the above procedure leads to determining an optimum image observing state, i.e., optimum set of the degree of retinal image's blur and the size of visual field, for viewing the presented image. Actually, the optimum image observing state is estimated by maximizing the evaluation function of Eq. (6) under Eq. (7) representing the interdependent relation between a blur parameter τ_1 and a visual field parameter τ_0 .

$$\begin{aligned} & \{\tau_0/2(\tau_0 - \tau_1)\} \log_{10}\{\tau_1/(2\tau_0 - \tau_1)\} + \\ & (1/2) \log_{10}\{(\tau_0 - \tau_1)^2/(2\tau_0\tau_1 - \tau_1^2)\} = \log_{10}(Cth^2/2) \end{aligned} \quad (7)$$

where $Cth^2/2$ means a constant threshold value. The evaluation function value is defined as a kind of impression strength received from the presented image, and it becomes large when the viewpoint overlaps with conspicuous positions such as edges of images.

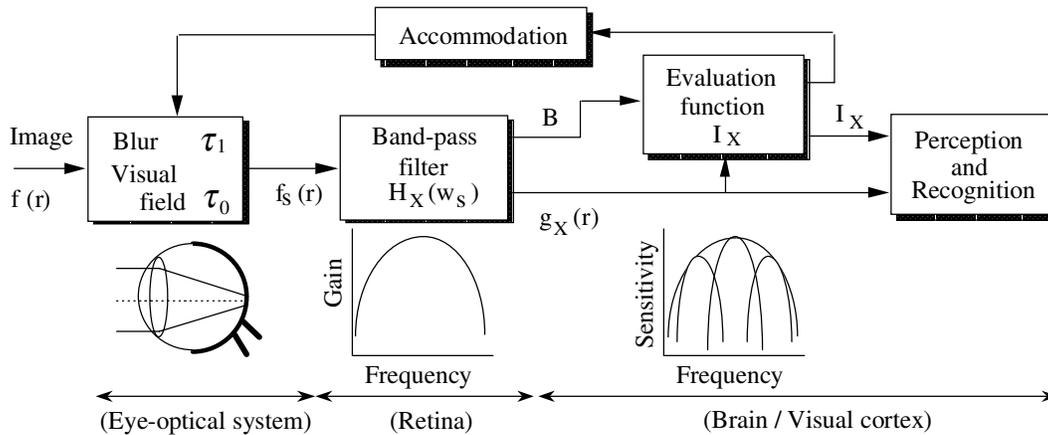


Figure 1. Block diagram of the cooperative human vision model.

One image observing state generates one narrow band-pass filter in the spatial frequency region. Consequently, as seen in the lower right end of Fig. 1, the model's spatial frequency characteristic is characterized by a multichannel structure just like in the conventional multichannel theory^{8,9}: it is composed of many narrow band-pass spatial channels, and the envelope of the narrow channels coincides with the experimentally measured sine wave's contrast sensitivity function. Estimating one optimum image observing state corresponds to selecting one of those channels. Therefore, the model can adaptively change its own image observing state or the position of its own narrow band-pass channel depending on the presented image's spatial frequency, even if the viewing distance is kept constant. The visual spatial frequency characteristic utilized in the conventional evaluation methods corresponds to the envelope of the multichannel structure. This is one of differences from the cooperative model. The center spatial frequency f_c (channel center frequency) of a narrow band-pass channel is connected as follows with the blur and visual field parameters.

$$f_c^2 = \log_e \{ (2\tau_o - \tau_i) / \tau_i \} / 16\pi^2 (\tau_o - \tau_i). \quad (8)$$

Reducing a channel center frequency leads to enlarging the retinal image's blur and expanding the visual field size, and vice versa.

It has already been shown that the model is effective in theoretically simulating how the human vision system changes its perceptual response characteristic depending on the property of presented images and observing conditions.²⁻⁴ In addition, since the model has its own viewpoint, it can reproduce the effect of subjects' viewpoints at subjective evaluation processes and detect positions where the image noise is conspicuous. Therefore, it is expected that noise evaluation metrics based on the model would agree well with human subjective judgments.

3. Model's Responses to Image Noise

3.1 Evaluated Images

Images utilized for evaluation were made by copying uniform gray patches of a test chart with three kinds of electrophotographic digital color copying machines, and were inputted to a computer by a color scanner with a resolution of 820 dpi (it corresponds to the resolution in the fovea centralis at a viewing distance of 30 cm) and a quantization number of 8 bits. The inputted images consisted of 256×256 pixels (7×7 mm in the copied images) and were converted into monochrome images on the computer. Since the area around the 256×256 pixel images was filled with a luminance level of 250, the images used for objective evaluation consisted of 512×512 pixels. The first digital color copying machine was a prototype model without optimization, and its halftone reproduction was based on a multilevel error diffusion method. The second and third ones were commercial products, and their halftone reproduction was based on a

digital screen method and a line screen method, respectively. The output resolution of all the machines was 600 dpi. The number of the evaluated images was 21: three kinds of test chart patch images different in optical density (OR0.1, OR0.2, OR0.4), twelve kinds of patch images (A0.1, A0.2, A0.4, B0.1, B0.2, B0.4, C0.1, C0.2, C0.4, D0.1, D0.2, D0.4) copied by the first machine under four kinds of conditions (A, B, C, and D), and six kinds of patch images (E0.1, E0.2, E0.4, F0.1, F0.2, F0.4) copied by the second and third machines. Among the four conditions, the electrophotographic process such as fixing or development was different, though the laser exposure system was identical. Figure 2 shows a test chart image (OR0.1) and six kinds of copied images (A0.1, B0.1, C0.1, D0.1, E0.1, F0.1). In addition, 21 standard patch images corresponding to 21 evaluated images were generated by the computer. Each image had the average gray level of the respective evaluated image.

3.2 Model's Response Characteristics

Figure 3 shows the model's responses calculated for a horizontal 1-dimensional image (512 pixels) taken out from A0.1 patch image as a function of viewpoint; (a), (b), and (c) represent the luminance level profile, the model's optimum channel center frequencies f_c , and the model's evaluation function values I_x , respectively; the hatched curves represent the model's responses calculated for the standard image.

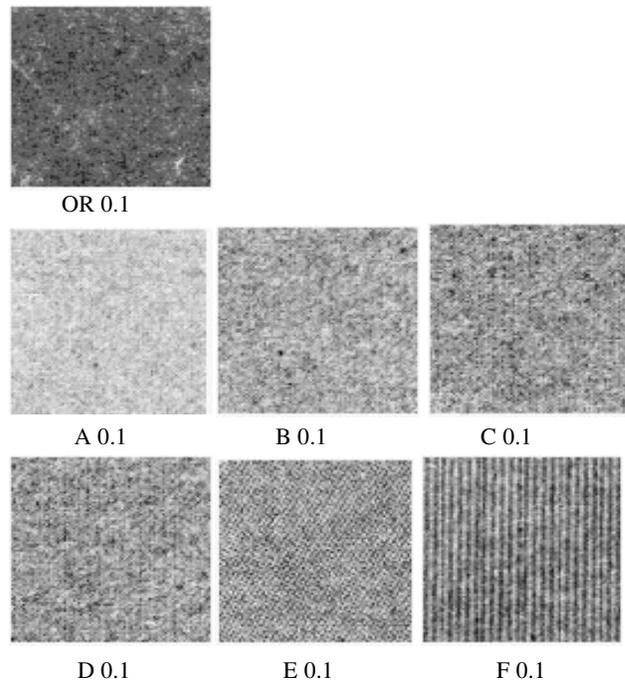


Figure 2. Examples of evaluated images. (density=0.1)

As seen from Fig. 3(b) and (c), the model's response was classified into two kinds of parts: in the first part, one response (fc_{1} , Ix_{1} or fc_{2} , Ix_{2}) was estimated at one viewpoint; in the second part, two kinds of responses (fc_{low} , Ix_{low} and fc_{high} , Ix_{high}) were estimated at one viewpoint. The existence of two kinds of responses in one viewpoint means that of two kinds of maxima in the evaluation function, in other words, that of two kinds of optimum image observing states suitable for observing the presented image at the viewpoint. The existence of two kinds of accommodative states at viewing an image has already been confirmed by measuring accommodative responses to multiple motion stimuli.¹⁰ The response values fc_{1} and Ix_{1} in the first part were quite different from ones for the standard image, while the response values fc_{2} and Ix_{2} were quite close to ones for the standard image. A decrease (or increase) in channel center frequency fc corresponds to an increase (or decrease) in retinal image's blur or a decrease (or increase) in presented image's spatial frequency, and the evaluation function value Ix is defined as a kind of impression strength or the degree of conspicuousness for presented images. Therefore, it can be understood that at the viewpoints of fc_{1} and Ix_{1} , the presented image includes high spatial frequency components which must be observed by reducing the retinal image's blur (fc_{1} is high), and that the components are considerably conspicuous (Ix_{1} is large); at the viewpoints of fc_{2} and Ix_{2} , the appearance for the noise is almost equal to that for the standard image. Seeing Fig. 3(a) from the above point of view, we can find a large difference in luminance level at the viewpoint where the evaluation function value Ix_{1} becomes a maximum (shown by an arrow). Actually, the difference in luminance level was perceived to be a nonuniformity in optical density. That is, it is concluded that the model's response values fc_{1} and Ix_{1} were caused by the presented image's nonuniformity.

The response values fc_{low} and Ix_{low} in the second part were quite close to ones for the standard image. This means that the appearance for the noise is almost equal to that for the standard image. In the other response, the value of fc_{high} was quite higher than that of fc_{1} , and the value of Ix_{high} was usually lower than that for the standard image. Therefore, it can be understood that in the case of fc_{high} and Ix_{high} , the presented image includes high spatial frequency components which must be observed by considerably reducing the retinal image's blur, and that the components are less conspicuous than the standard image. Seeing Fig. 3(a) at these viewpoints, we can find a periodically fluctuating component. The component was dot structure noise or modulation noise. That is, it is concluded that the dot structure noise caused two kinds of model's responses. Generally, a lot of attention is required to make the retinal image's blur less and make the accommodative state closer to the just-focused condition. Therefore, an interpretation in the second part is that the appearance for the noise is almost equal to that for the standard image if the attention is not concentrated (fc_{low}

and Ix_{low}), while the dot structure noise can be perceived if the attention is concentrated (fc_{high} and Ix_{high}). However, the dot structure noise is interpreted as being relatively inconspicuous because a lot of attention is required to perceive it.

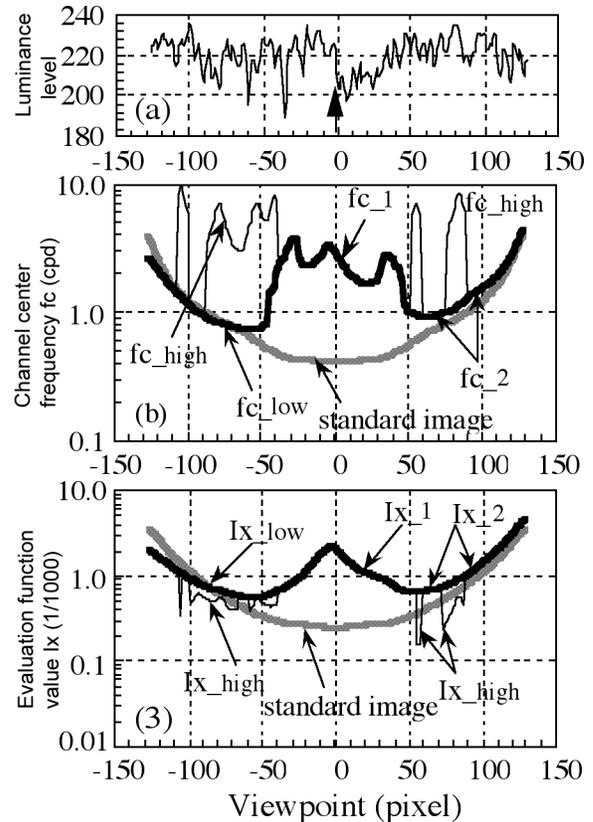


Figure 3. Model's responses for a horizontal line image of A0.1.

4. Vision-Based Image Noise Evaluation

4.1 Image Noise Evaluation Metrics

Based on the model's responses, a mean-square-log (MSL) error per pixel of the evaluation function value was chosen as a metric for image noise. Fig. 4 shows the MSL value (dB^2) per pixel as a parameter of the optical density (0.1, 0.2, and 0.4); (a) represents the degree of conspicuousness for the nonuniformity in optical density in comparison with the standard image, and was calculated by summing up the square of the difference in logarithmic value between the fat solid line and the hatched line in Fig. 3(c) in all viewpoints and dividing it by the number of all viewpoints, although, if Ix_{high} was larger than Ix_{low} , it was regarded as a part of the fat solid line; (b) represents the degree of conspicuousness for the dot structure noise, and was calculated by summing up the square of the difference in logarithmic value between Ix_{low} and Ix_{high} and dividing it by the number of viewpoints where two kinds of optimum image observing states were estimated.

From Fig.4(a) the following results were derived: (1) the difference in nonuniformity in density was very small among images A, B, C, and D; (2) the degree of nonuniformity in images E and F was clearly smaller than that in images A, B, C, and D, and that in image F was the least except for the test chart image OR; (3) the degree of nonuniformity decreased with an increase in optical density. On the other hand, a decrease in I_{x_high} , i.e., an increase in the MSL value in Fig. 4(b), means a decrease in the dot structure noise. Judging from this point of view, from Fig.4(b) the following results were derived: (1) the difference in dot structure noise was very small among images A, B, C, and D; (2) the degree of dot structure noise in images E and F was clearly smaller than that in images A, B, C, and D, and that in image F was the least except for the test chart image OR; (3) the degree of dot structure noise decreased with an increase in optical density. The above results mean that images E and F are clearly more excellent in image quality than images A, B, C, and D, and especially the quality of image F is most excellent except for the test chart image OR.

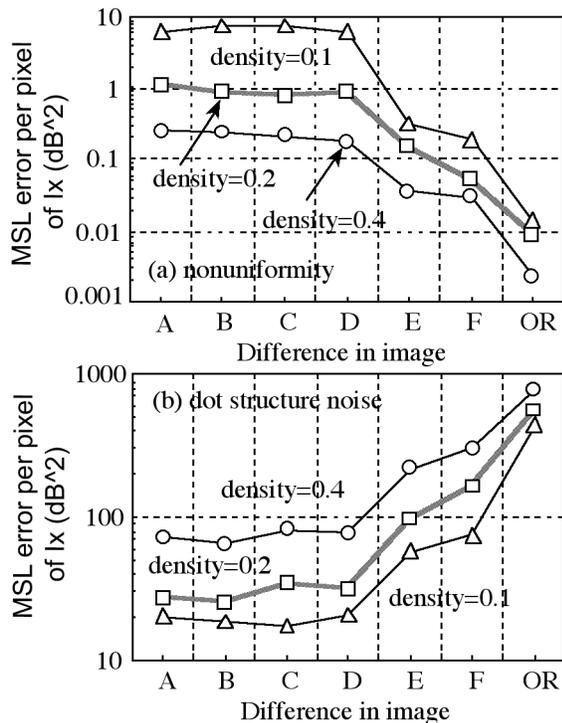


Figure 4. MSL error values per pixel of the model's evaluation function value I_x .

4.2 Subjective Evaluation

A subjective evaluation experiment was carried out using Scheffe's method of paired comparison in order to verify the objective evaluation results. Images utilized for subjective evaluation were 21 patch images outputted from the digital color copying machines. Subjects were 9 students without experiences in subjective evaluation, and were instructed to evaluate the degree of image noise or

granularity for all pairs of 7 images (A, B, C, D, E, F, and OR) in each optical density by using a method of 7 categories (3: very good, 2: quite good, 1: a little good, 0: equal, -1: a little bad, -2: quite bad, -3: very bad).

Fig. 5 shows the relationship between the MSL error values for 7 images (A, B, C, D, E, F, and OR) and the subjective evaluation values (psychological image noise scores) as a parameter of optical density; the MSL error values used in (a) and (b) represent the degree of conspicousness for the nonuniformity in optical density (Fig. 4(a)) and the dot structure noise (Fig. 4(b)), respectively; the numeral in the parentheses means the derived correlation coefficient. As seen from Fig. 5, the correlation coefficient between the MSL error values and the subjective scores was from 0.985 to 0.999. The above result means that the proposed two kinds of MSL error values correlate closely with the psychological image noise score and that they can be fully utilized as metrics for image noise or granularity.

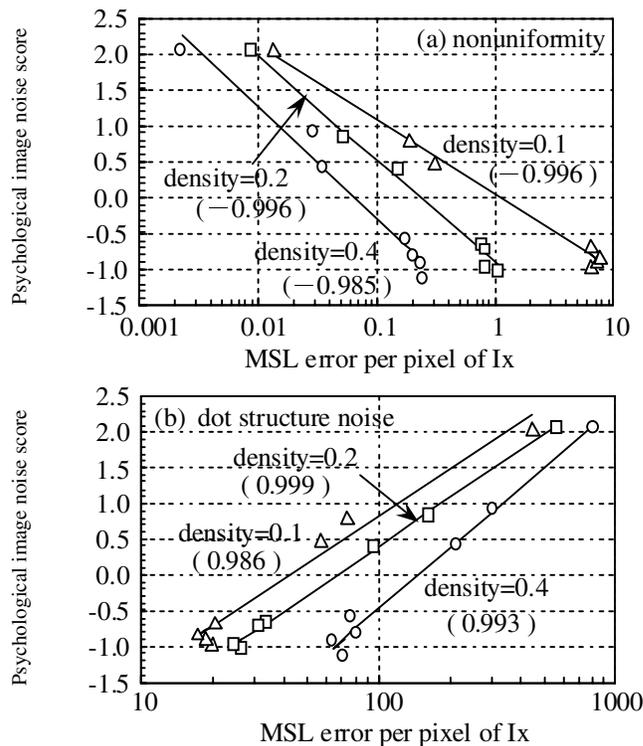


Figure 5. Relationship between MSL error per pixel of I_x and psychological image noise score.

4.3 Evaluation for Color Registration Noise

Fig. 6 shows the MSL value (dB^2) per pixel for 7 color patch images (Yellow, Magenta, Cyan, Red, Green, Blue, Black) at an optical density of 0.1, and was calculated in the same way as Fig. 4; (a) represents the degree of conspicousness for the nonuniformity in optical density in comparison with the standard image; (b) represents the degree of conspicousness for the dot structure noise. In

the actual calculation, all the color patch images were converted into monochrome images on the computer, and each monochrome image was utilized as an input to the cooperative model. From Fig.6 the following results were derived: (1) the amount of image noise in the single-colored patches (Y, M, and C) was larger than that of the other color patches; (2) the amount of image noise tended to decrease every time the color printing process was repeated. Therefore, it can be concluded that the performance of the digital color copying machines used here is strongly influenced by the quality level of the single-colored patches. The above results suggest that it is important to improve the quality level of the single-colored images, especially yellow image, to enhance the performance of the digital color copying machines used here more and more.

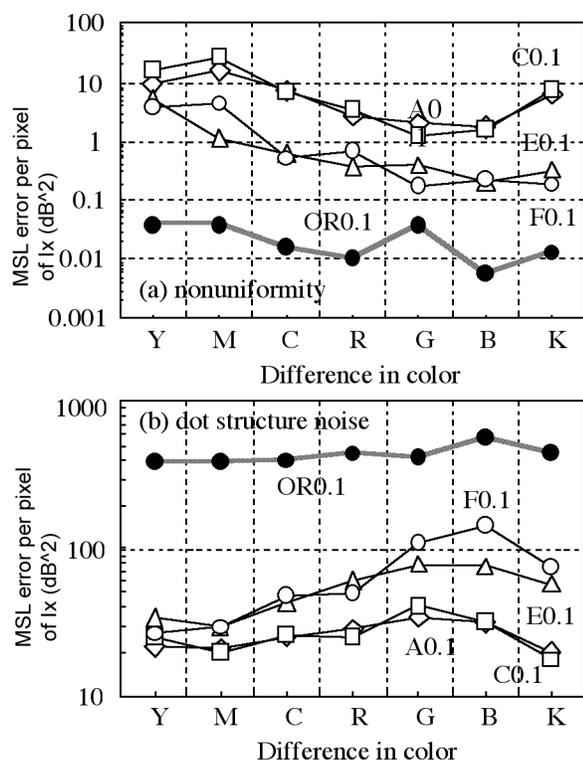


Figure 6. MSL value per pixel for 7 color patch images (Yellow, Magenta, Cyan, Red, Green, Blue, Black) at an optical density of 0.1.

5. Conclusion

A new objective image noise evaluation method using a cooperative human vision model has been proposed. In order to examine its effectiveness, it was applied to noise evaluation for halftone patch images with a uniform density outputted from three kinds of digital color copying machines. The following results were obtained: (1) the vision model simulated correctly human vision's perceptual responses to the patch images' noise at every viewpoint; (2) two kinds of evaluation criteria (nonuniformity in optical density and dot structure noise) were constructed based on the model's responses; (3) the above two values correlated closely with the psychological image noise score based on Scheffe's method of paired comparison (correlation coefficient > 0.96); (4) the quality level of the single-colored images was a factor to affect the performance of the copying machines used. These results suggest that the proposed method is effective in producing objective evaluation results in good agreement on subjective judgments of humans and that human vision's adaptively-changing perceptual responses are appropriately reflected in the metrics.

References

1. R.P. Dooly et.al, *J. Appl. Photogr. Eng.*, **5**, 190 (1979).
2. T. Matsui, *Trans. IEICE*, (in Japanese) **J71-D**, 2669 (1988).
3. T. Matsui and S. Hirahara, *Trans. IEICE*, (in Japanese) **J75-DII**, 1597 (1992).
4. T. Matsui, *Trans. IEICE*, (in Japanese) **J79-DII**, 889 (1996).
5. K. Kasai et.al, *Trans. IECE*, (in Japanese) **57-D**, 261 (1974).
6. W.N. Charman et.al, *Vision Research*, **17**, 129 (1977).
7. D.A. Owen, *Vision Research*, **20**, 159-167 (1980).
8. F.W. Campbell et.al, *J. Physiol.*, **197**, 551 (1968).
9. C. Blakemore et.al, *J. Physiol.*, **203**, 237 (1969).
10. T. Matsui, *J. ITE*, (in Japanese) **55**, 1460 (2001).

Biography

Toshikazu Matsui received his B.S., M.E., and Ph.D. degrees in Electrical Engineering from Waseda University, Japan, in 1977, 1979, and 1997, respectively. He joined Research & Development Center of Toshiba Corporation in 1980. From 1994 to 1996, he had been a Senior Researcher at ATR Human Information Processing Research Laboratories. Since 1998 he has been an associate professor in Graduate School of Engineering, Gunma University, Japan. His work has primarily focused on human vision, visual signal processing, and image quality evaluation.