PAPER

Video Quality Assessment using Spatio-Velocity Contrast Sensitivity Function

Keita HIRAI[†a)], **Jambal TUMURTOGOO**[†], ***Student Members***, **Ayano KIKUCHI**[†], ***Nonmember***, **Norimichi TSUMURA**[†], **Toshiya NAKAGUCHI**[†], ***and*** **Yoichi MIYAKE**[††], ***Members***

SUMMARY Due to the development and popularization of high-definition televisions, digital video cameras, Blu-ray discs, digital broadcasting, IP television and so on, it plays an important role to identify and quantify video quality degradations. In this paper, we propose SV-CIELAB which is an objective video quality assessment (VQA) method using a spatio-velocity contrast sensitivity function (SV-CSF). In SV-CIELAB, motion information in videos is effectively utilized for filtering unnecessary information in the spatial frequency domain. As the filter to apply videos, we used the SV-CSF. It is a modulation transfer function of the human visual system, and consists of the relationship among contrast sensitivities, spatial frequencies and velocities of perceived stimuli. In the filtering process, the SV-CSF cannot be directly applied in the spatial frequency domain because spatial coordinate information is required when using velocity information. For filtering by the SV-CSF, we obtain video frames separated in spatial frequency domain. By using velocity information, the separated frames with limited spatial frequencies are weighted by contrast sensitivities in the SV-CSF model. In SV-CIELAB, the criteria are obtained by calculating image differences between filtered original and distorted videos. For the validation of SV-CIELAB, subjective evaluation experiments were conducted. The subjective experimental results were compared with SV-CIELAB and the conventional VQA methods such as CIELAB color difference, Spatial-CIELAB, signal to noise ratio and so on. From the experimental results, it was shown that SV-CIELAB is a more efficient VQA method than the conventional methods.
key words: *objective video quality assessment, spatio-velocity contrast sensitivity function, image diffrence, subjective evaluation, rank order correlation coefficient*

1. Introduction

Recently, due to the development and popularization of high-definition televisions, digital video cameras, Blu-ray discs, digital broadcasting, IP television and so on, high-quality videos have been widely used in our life. As high-quality videos become popular, it plays an important role to identify and quantify video quality degradations due to video data compression.

Since humans are the ultimate evaluators of video quality, the most reliable way of video quality assessments (VQAs) is a subjective evaluation method. However the cost of a subjective evaluation is expensive and it is not an appropriate way for a versatile VQA method. On the other hand, an objective VQA method is not expensive compared with subjective methods. In general, factors of image quality in objective evaluations are quantified by the criteria such as sharpness, color reproduction, tone reproduction and noise characteristics. The simplest and most widely used image quality assessment (IQA) method is the mean square error (MSE) or peak signal to noise ratio (PSNR). But they are not very well correlated with perceived visual quality [1-5]. Since images are perceived through the human visual system, objective IQA or VQA methods should be designed by incorporating human visual characteristics [1]. Based on this observation, Wang et al. proposed structural similarity (SSIM) [6-8]. Though SSIM is an IQA method based on the assumption of human visual perception, it does not contain exact human visual characteristics.

In conventional IQA or VQA methods with the human visual characteristics, contrast sensitivity functions (CSFs) are frequently used. CSFs are a modulation transfer function of the human visual system, and have band-pass characteristic in spatial frequency domain [9]. In 1996, Zhang et al. proposed Spatial-CIELAB (S-CIELAB) [10]. To compute image degradations between original and distorted images, they applied a CSF in filtering images. The criteria can be obtained by calculating CIELAB color differences between filtered original and distorted images at each pixel. Though S-CIELAB is useful for still image quality assessment, temporal characteristics in human visual system were not considered and it is not enough for VQA. To evaluate video quality, Tong et al, proposed spatio-temporal CIELAB (ST-CIELAB), which is an extension of S-CIELAB [11]. They used a spatio-temporal CSF [12][13] to filter original and distorted videos. Though ST-CIELAB was proposed as a VQA method, a previous research has reported that ST-CIELAB is not so much effective compared to S-CIELAB [14]. This reason is that ST-CSF does not contain eye movement characteristics. When observers evaluate video quality, it is considered that their eyes track moving objects in the video. To build more useful VQA methods, the eye movement characteristics should be incorporated.

[†]Graduate School of Advanced Integration Science, Chiba University, 1-33, Yayoi-cho, Inage-ku, Chiba, 263-8522, Japan
[††]Research Center for Frontier Medical Engineering, Chiba University, 1-33, Yayoi-cho, Inage-ku, Chiba, 263-8522, Japan
a) E-mail: hirai@graduate.chiba-u.jp
DOI: 10.1587/transinf.E0.D.1

In this paper, therefore, we propose SV-CIELAB which is a VQA method using a spatio-velocity contrast sensitivity function (SV-CSF) [15-17]. The SV-CSF consists of the relationship among contrast sensitivities, spatial frequencies and velocities of stimuli. It also contains eye movement characteristics that follow moving objects. We assumed that eyes track moving objects in videos when observers evaluate the video quality. Based on this observation, the SV-CSF was applied to filter original and distorted videos. The criteria in our method are obtained by calculating image differences between filtered videos. Furthermore we conducted subjective experiments for validating SV-CIELAB. The subjective results were compared with SV-CIELAB and the conventional VQA methods which are PSNR, SSIM, CIELAB color difference, S-CIELAB and ST-CIELAB.

2. Related Work

2.1 Spatial-CIELAB and Spatio-Temporal CIELAB

Figure 1 shows the overview of the color difference calculations in S-CIELAB or ST-CIELAB [10-14]. Original R, G, B images and distorted R, G, B images are respectively transformed to the opponent color components, A (luminance channel), T (r/g channel), D (b/y channel). Then, the spatial frequency filtering in S-CIELAB and the spatio-temporal frequency filtering in ST-CIELAB are performed. The filters are CSFs which represent band-pass characteristics in human visual system. Those filtered images are transformed to X, Y, Z colorimetric values and then L*, a*, b* values. Finally color difference is calculated pixel-by-pixel and then the mean difference is calculated.

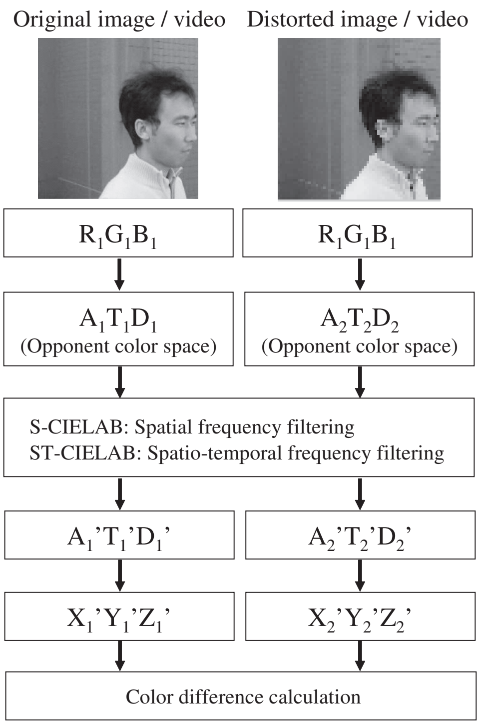


Fig. 1 Overview of S-CIELAB and ST-CIELAB.

2.2 Spatio-Velocity Contrast Sensitivity Function

The SV-CSF model was originally proposed by S. Daly et al. [15][16]. They proposed the SV-CSF model based on the experimental data of some previous works [9][12] and their brief experiments. However the parameters in the SV-CSF model were not validated sufficiently. In 2007, more detailed experiments to measure the contrast sensitivities were conducted by Hirai et al [17]. In the experiments, Gabor patterns were used as the stimuli which moved from left to right on a display device. A fixation point was also displayed at the center of the stimuli and the point moved along the motion of the stimuli. The observers were instructed to fix their gazes on the fixation point, and contrast sensitivities were measured. Based on the measured data, they optimized the parameters in the SV-CSF model.

Figure 2 represents the SV-CSF model [17]. The SV-CSF consists of the relationship among contrast sensitivities, spatial frequencies and velocities of stimuli. As shown in Fig.2, the SV-CSF has band-pass characteristics and the peak of contrast sensitivities with 0 degrees/second (degree means visual degree) is around 3 cycles/degree. As the velocity increases, the peak becomes close to lower frequency. In the experiments to measure contrast sensitivities, ten subjects participated [17]. The results of each subject are same tendency which indicates that SV-CSF has band-pass characteristics and the peak becomes close to lower frequency as the velocity of stimulus increase.

As shown in Fig.2, the contrast sensitivities depend on the velocity of the moving stimuli. If eyes follow the fixation point perfectly, the retinal image when looking at moving stimuli are ideally the same as that when looking at still stimuli. However, the contrast sensitiv-

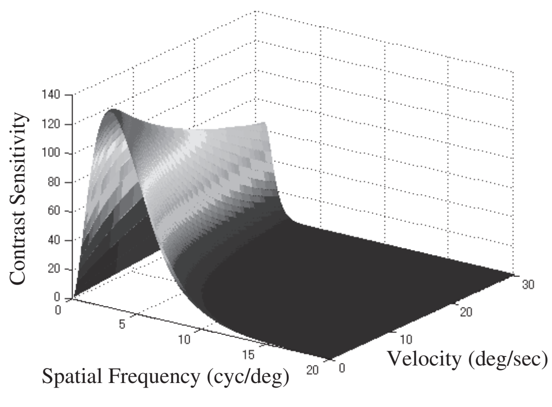


Fig. 2 SV-CSF model.

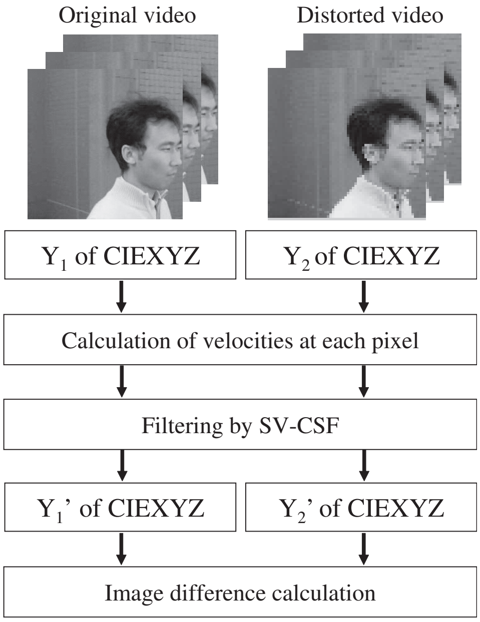


Fig. 3 Overview of the proposed SV-CIELAB.

ities changed depending on the velocity of the moving stimuli. S. Daly described that this reason is caused by smooth pursuit eye movements [15]. The smooth pursuit eye movements reduce the accuracy of tracking moving stimuli. In other words, the smooth pursuit eye movements change the characteristics of the contrast sensitivities when looking at moving stimuli.

3. SV-CIELAB

3.1 Overview

SV-CIELAB is a VQA method using the SV-CSF. However the SV-CSF model was proposed for the only luminance channel in the human visual system. Therefore, in this research, gray-scale videos are addressed. Y values (luminance channel) of CIEXYZ color space are used in the processing of the videos. Figure 3 shows the overview of the proposed SV-CIELAB. As described above, the SV-CSF model contains velocity axis. Therefore, first, the velocities at each pixel are acquired. To obtain the velocities, we calculated optical flows by using the Bergen's method [18]. Next, the original and distorted videos are filtered using the optical flows and the SV-CSF. Finally, the criteria in SV-CIELAB are obtained by calculating image differences between filtered original and distorted videos.

3.2 Filtering in SV-CIELAB

Figure 4 shows the filtering process using the SV-CSF. In S-CIELAB and ST-CIELAB, input images or videos are filtered in spatial or temporal frequency domain.

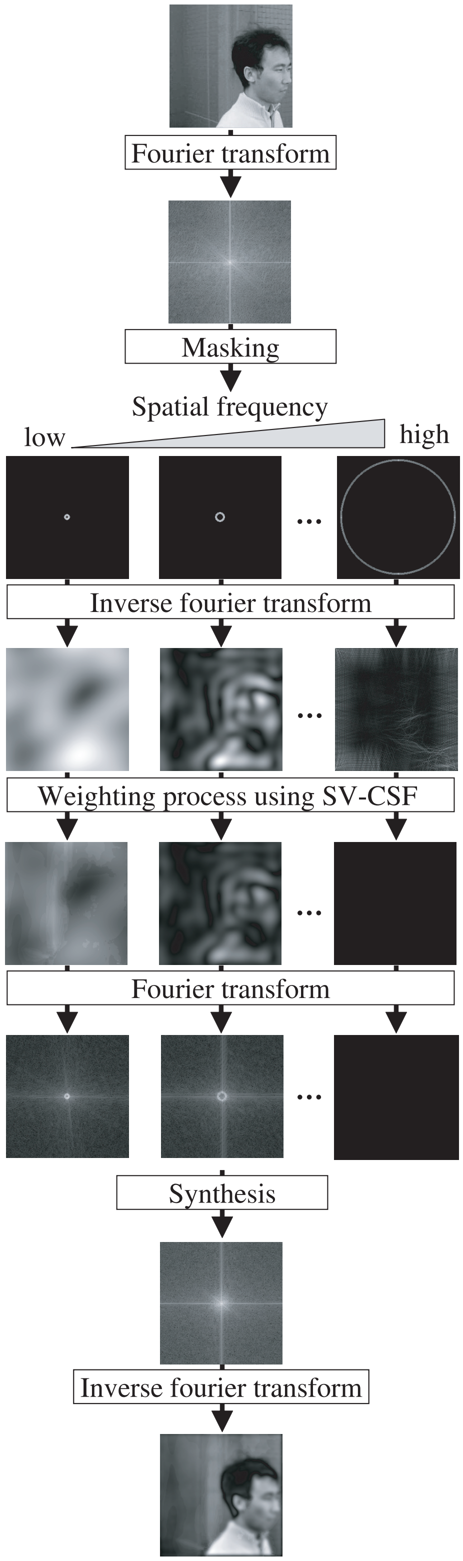


Fig. 4 Filtering process in SV-CIELAB.

However, the SV-CSF cannot be applied in the frequency domains because the spatial coordinate information is required when using velocity information at each pixel. Therefore, in filtering by the SV-CSF, we

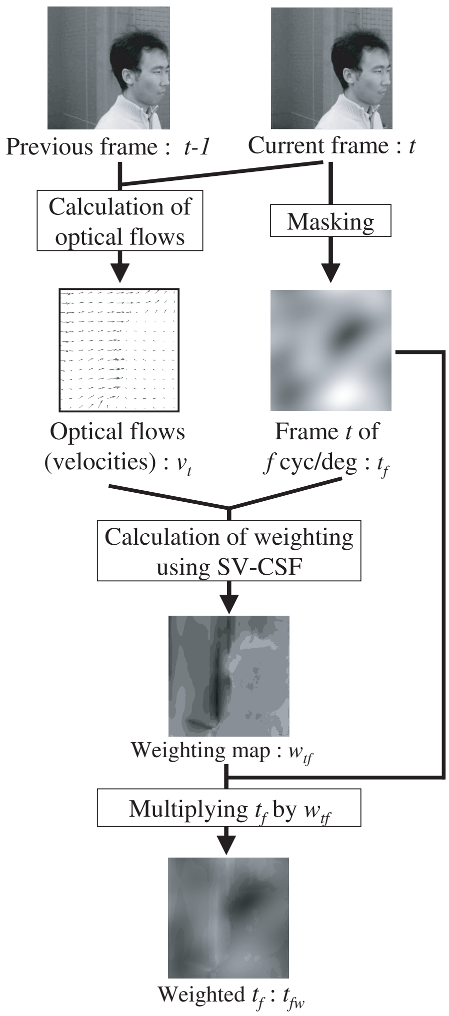


Fig. 5 Weigting process using SV-CSF.

obtain video frames separated in spatial frequency domain. Each separated frame has the information of one cyc/deg. By using velocity information, the frames separated by each spatial frequency are weighted by contrast sensitivities in the SV-CSF model (described in Section 3.3). A final filtered frame is obtained by synthesizing the weighted frames.

3.3 Weighting Map

Figure 5 shows the weighting process using the SV-CSF. In this process, weighting map w at frame t configured by specific spatial frequency f (cyc/deg) is generated by a following equation.

$$w_{tf}(x,y) = \text{SV-CSF}_{nor}(f, v_t(x,y)) \tag{1}$$

where x and y are spatial coordinates, v_t is velocity (deg/sec) at frame t. SV-CSF$_{nor}$ is the normalized SV-CSF which range is from 0 to 1. f is given by the frames separated by each spatial frequency as described in Section 3.2. The SV-CSF [17] is computed by

$$\begin{aligned} &\text{SV-CSF}(f,v) = \\ &\quad kc_0c_1c_2(v+c_v)(c_1 2\pi f)^2 exp\left(-\frac{c_1 4\pi f}{f_{max}}\right) \\ &k = 6.1 + 7.3|log(c_2(v+c_v)/3)|^3 \\ &f_{max} = 45.9/(c_2(v+c_v)+2) \end{aligned} \tag{2}$$

where c_0, c_1, c_2 and c_v are parameters : $c_0 = 1.00$, $c_1 = 0.56$, $c_2 = 0.48$ and $c_v = 5.1$. Finally, the weighted frame t of f cyc/deg is calculated by

$$t_{fw}(x,y) = w_{tf}(x,y)t_f(x,y) \tag{3}$$

3.4 Image Difference Calculation

The criteria in SV-CIELAB are obtained by calculating image differences between filtered original and distorted videos. To calculate the differences, the filtered videos are transformed to L*, a*, b* values in CIELAB color space. The image difference is computed pixel-by-pixel and then the mean difference is obtained as follow.

$$\text{Image difference} = \frac{\sum_{x,y,t}\sqrt{(L_o(x,y,t) - L_d(x,y,t))^2}}{N_x N_y N_t} \tag{4}$$

where N_x, N_y and N_t are the number of samples in x, y and t, respectively. L_o and L_d are L* (lightness) values of the filtered original and distorted video.

4. Validation

In general, VQA methods are required to address video quality factors such as sharpness, tone reproduction, color reproduction and noise characteristics. In particular, noise such as random noise, mosquito noise and blockiness often occur due to broadcasting or video data compression, and it is important to evaluate degradation caused by the noise. Therefore, as the first step, distorted videos generated by adding random noise were used to validate SV-CIELAB.

For the validation of SV-CIELAB, two kinds of subjective evaluation experiments were conducted. In the validation, subjective experimental results were compared with SV-CIELAB and the conventional VQA methods which are PSNR, SSIM, CIELAB color difference, S-CIELAB and ST-CIELAB.

In the calculation of the objective scores, CIELAB color difference, S-CIELAB and ST-CIELAB were obtained by the ways described in section 2.2 and Eq.4. PSNR is defined as

$$\begin{aligned} &\text{PSNR} = 10\log_{10}\left(\frac{{I_{max}}^2}{\text{MSE}}\right) \\ &\text{MSE} = \frac{1}{N_x N_y N_t}\sum_{x,y,t}(I_o(x,y,t) - I_d(x,y,t))^2 \end{aligned} \tag{5}$$

where x and y are spatial coordinates and t is frame number. N_x, N_y and N_t are the number of samples in

x, y and t, respectively. I_o and I_d are pixel values of the filtered original and distorted video. I_{max} is a constant which represents the dynamic range of pixel intensities (e.g., for 8 bits/pixel grayscale image, I_{max}=255). SSIM [6] was also calculated as one of typical conventional VQA methods. SSIM is defined as

$$\mathrm{SSIM}(t) = \frac{(2\mu_o(t)\mu_d(t) + C_1)(2\sigma_{od}(t) + C_2)}{(\mu_o(t)^2 + \mu_d(t)^2 + C_1)(\sigma_o(t)^2 + \sigma_d(t)^2 + C_2)} \quad (6)$$

where μ_o, μ_d and σ_o, σ_d are the mean and standard deviation of the frame t in the original and the distorted videos, respectively. σ_{od} is the cross correlation between the mean-removed frame t in the original video and that in the distorted video, and C_1 and C_2 are two constants : $C_1 = 6.5$ and $C_2 = 58.5$. SSIM is a value between -1 and 1, and value 1 means that the quality of an original frame is the same as the one of a distorted frame. Finally, SSIM is a simple average over all frames.

(a) (b)

Fig. 6 Displayed videos. (a) Sample 1 : forest. (b) Sample 2 : standing men.

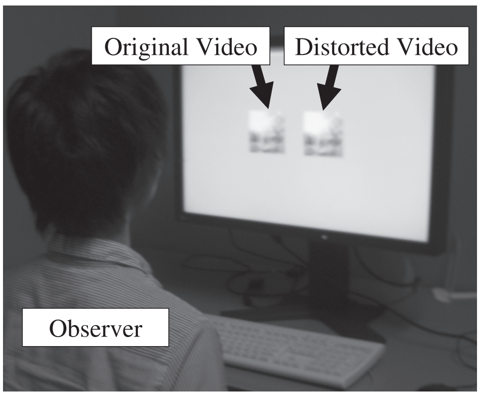


Fig. 7 Experimental room.

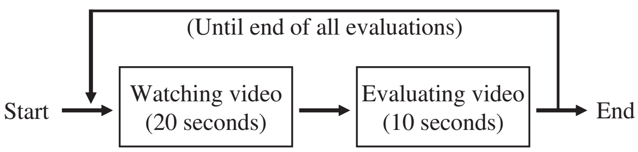


Fig. 8 Experimental procedure.

Table 1 Setups of display device.

type	LCD
size	27 inches
pixels	1980 × 1280 pixels
pixel pitch	0.303 mm
maximum luminance	340 cd/m^2
background luminance around videos	170 cd/m^2 (half of maximum luminance)
gamma	2.2
color temperature	6500 ^{o}C

4.1 Experiment A

In Experiment A, assuming that the video contents of forest and men are commonly used, we prepared two videos as shown in Fig. 6 (Sample 1: forest, and Sample 2: standing men). The velocities and directions of motion in each pixel of the videos were same because the motions of the videos were generated by virtual panning of a camera. The velocities in the videos were 0, 2.5, 5 and 10 deg/sec and the directions are horizontal scrolls. The distorted videos were generated by adding random noise of 0, 7.5, 10 and 12.5% (maximum luminance levels of random noise are 7.5, 10 and 12.5% of 340cd/m^2 which is the maximum luminance of the LCD). The random noise was generated frame-by-frame based on general random noise in broadcasting. Totally 32 videos were evaluated. The videos are 30 frames/second, the size is 256 × 256 pixels, and the time of each video is 10 seconds.

Figure 7 shows the experimental room. The setup of illumination in the room is dark condition. Fifteen observers participated in Experiment A. In the VQA methods, the objective scores are obtained by calculating the difference between the original and distorted videos. Therefore, in the subjective experiments, the original and distorted videos were displayed at the same time and observers were instructed to evaluate the difference between the videos. The observers were not instructed how to watch the videos. In other words, the observers could watch the original and distorted videos freely while the videos were displayed. The viewing condition in this instruction way is close to the natural viewing condition while viewing videos, and it was used in the conventional evaluation of S-CIELAB or ST-CIELAB [14]. In these experimental conditions, we assumed that eyes tracked moving objects in videos while observers evaluated the video quality. We also considered that the eyes could track moving objects while the observers were watching the videos, because similar experimental methods were used to evaluate quality of moving images which were displayed on LCDs [19, 20]. In Experiment A, the observers evaluated the degraded level by the scale of 1 to 5 (1: the same, 2: slightly different, 3: different, 4: definitely different and 5: very different). Table 1 shows the setups of the display de-

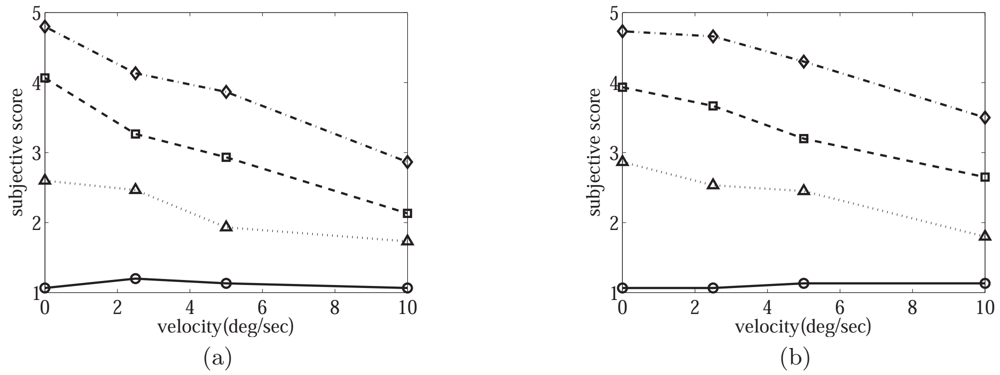


Fig. 9 Relationships between subjective scores and velocities of videos. Noise levels are ○:0%, △:7.5%, □:10% and ◇:12.5%. (a) Sample 1 : forest. (b) Sample 2 : standing men.

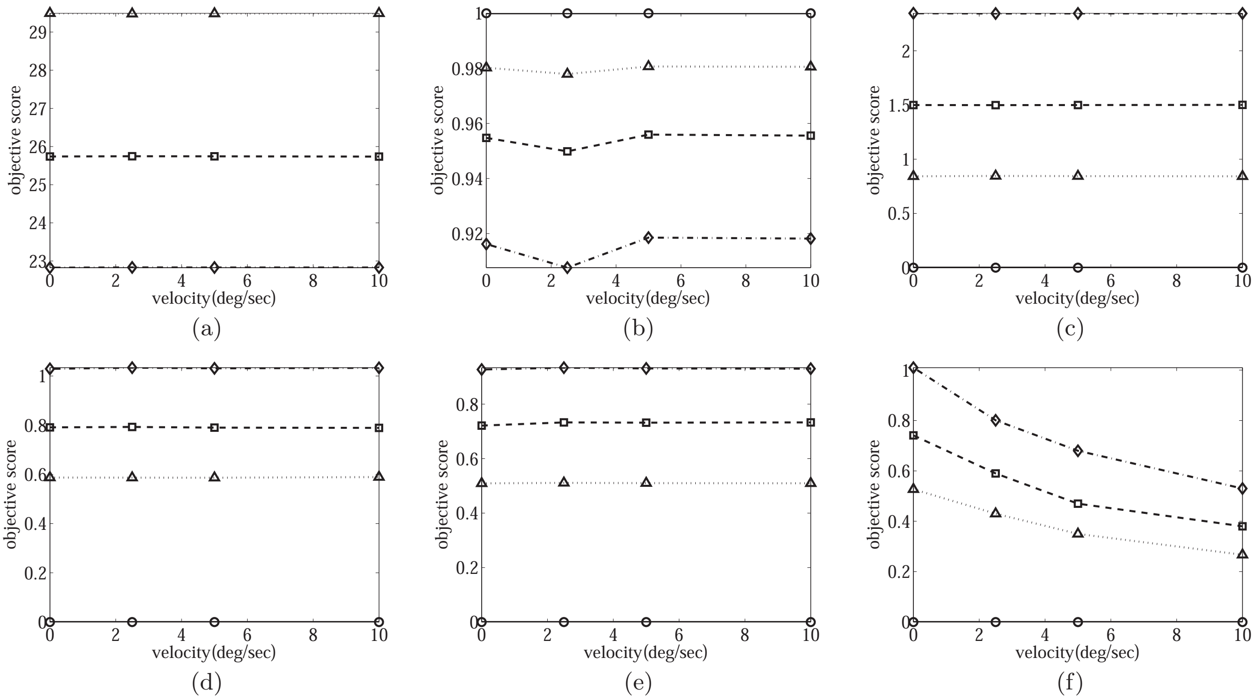


Fig. 10 Relationships between objective scores and velocities in Sample 1. Noise levels are ○:0%, △:7.5%, □:10% and ◇:12.5%. (a) PSNR. (b) SSIM. (c) CIELAB. (d) S-CIELAB. (e) ST-CIELAB. (f) SV-CIELAB.

vice. We prepared a 27 ” LCD in the experiment. The viewing distance is 400mm which are decided based on the standard viewing distance recommended by ITU [21]. Figure 8 shows the experimental procedure. First, observers watched a video for 20 seconds. The video of 20 seconds were presented by running a video of 10 seconds continuously. By such displaying procedure, we consider it is possible to evaluate same areas in the original and distorted video easily. After watching the video, observers evaluated the degraded level in 10 seconds. In this evaluating period, a gray image was displayed. The luminance of the gray image is the same as the one of the background shown in Table 1. The distorted videos were displayed in random order, and the watching and evaluating period were repeated until all of the degraded videos have been evaluated.

Figure 9 shows the results of the relationships between the subjective evaluation scores and the velocities of the videos. As the velocities increase, the subjective scores become lower values. In other words, the observers cannot perceive additive noise accurately in the videos with high velocities.

Figure 10 and 11 show the results that describe the relationships between the objective evaluated scores and the velocities in the videos (Sample 1 and 2). In the results of PSNR (Fig. 10(a) and Fig. 11(a)), the plots with 0% noise are excepted because the PSNR between the original videos and the distorted video of 0% noise

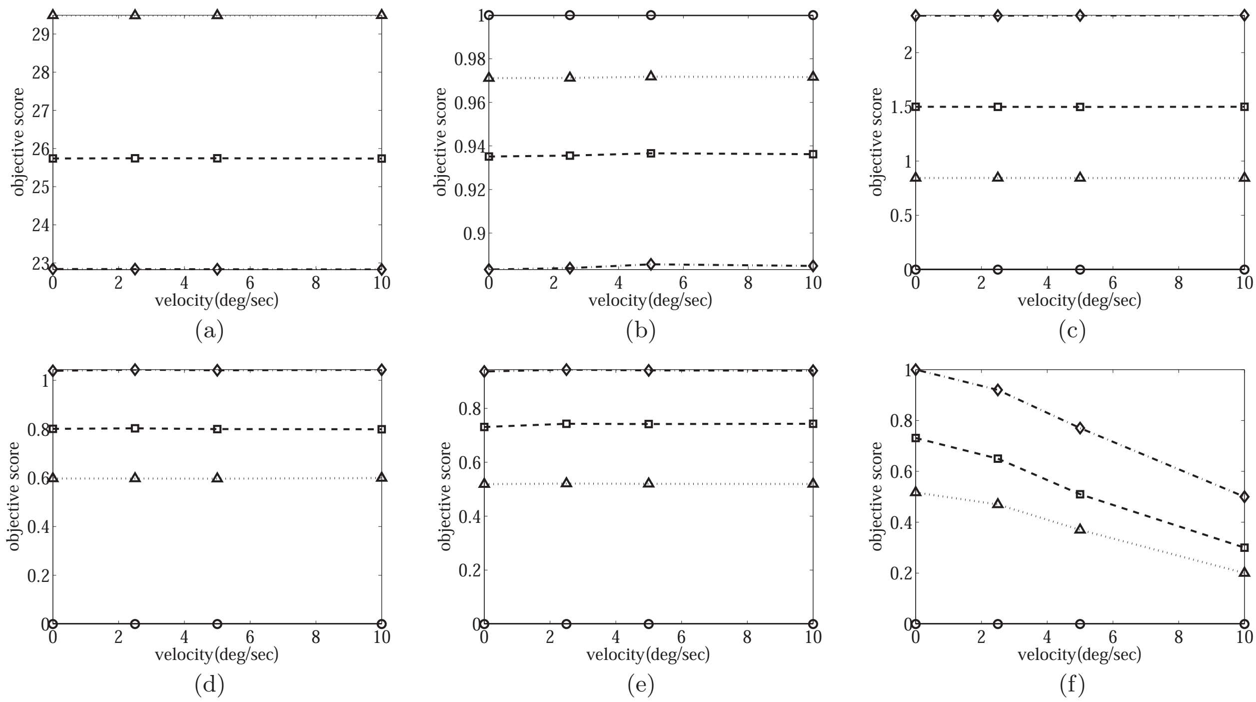


Fig. 11 Relationships between objective scores and velocities in Sample 2. Noise levels are ○:0%, △:7.5%, □:10% and ◇:12.5%. (a) PSNR. (b) SSIM. (c) CIELAB. (d) S-CIELAB. (e) ST-CIELAB. (f) SV-CIELAB.

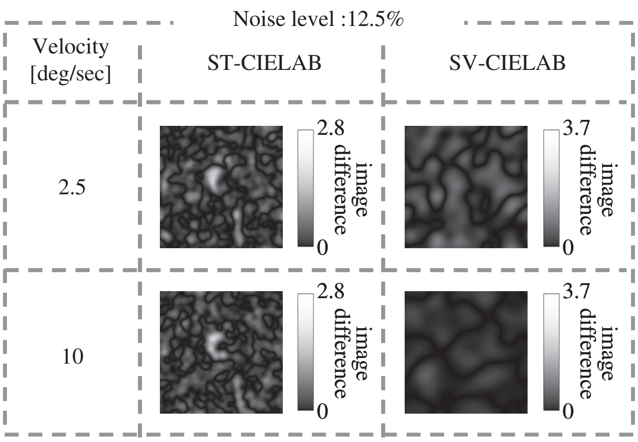


Fig. 12 Image differences between an original and distorted frame in Sample 1.

cannot be calculated. The objective scores of PSNR, SSIM, CIELAB, S-CIIELAB and ST-CIELAB shown in Fig. 10 and 11 are approximately the same values in each noise level respectively. These results mean that the conventional methods practically are not affected by the changes of the velocities. On the other hand, the objective scores of SV-CIELAB become lower values as the velocities increase. The tendency of SV-CIELAB is similar to the subjective results shown in Fig. 9. Figure 12 shows the images differences between an original and distorted frame in Sample 1. The figures also show that ST-CIELAB cannot take into account the change of the velocities whereas the objective scores of SV-CIELAB become lower values with higher velocities.

Figure 13 shows the relationships between subjective and objective scores. In PSNR and SSIM, the image qualities of the distorted videos are better as the objective scores are higher. On the other hand, in CIELAB color difference, S-CIELAB, ST-CIELAB and SV-CIELAB, the image qualities are worse as the ones are higher. Therefore, in Fig. 13(a) and (b), the values of the horizontal axes are the reciprocals of PSNR and SSIM to conform to the CIELAB color difference, S-CIELAB, ST-CIELAB and SV-CIELAB. Table 2 also shows the Spearman ' s rank order correlation coefficients (ROCC) [22] between subjective scores and the proposed and conventional VQA methods. ROCC is given by

$$\mathrm{ROCC} = 1 - \frac{6\sum d^2}{n(n^2-1)} \tag{7}$$

where n is the number of evaluated videos in the experiment and d is the difference between the ranks in subjective and objective scores. ROCC is one of the metrics for the evaluation of video quality measures [23]. Its advantage is in its robustness because it is independent of any fitting functions that attempt to find a nonlinear mapping between the objective and the subjective scores. From the results shown in Fig. 13 and Table 2, SV-CIELAB is the more efficient VQA method than the conventional methods.

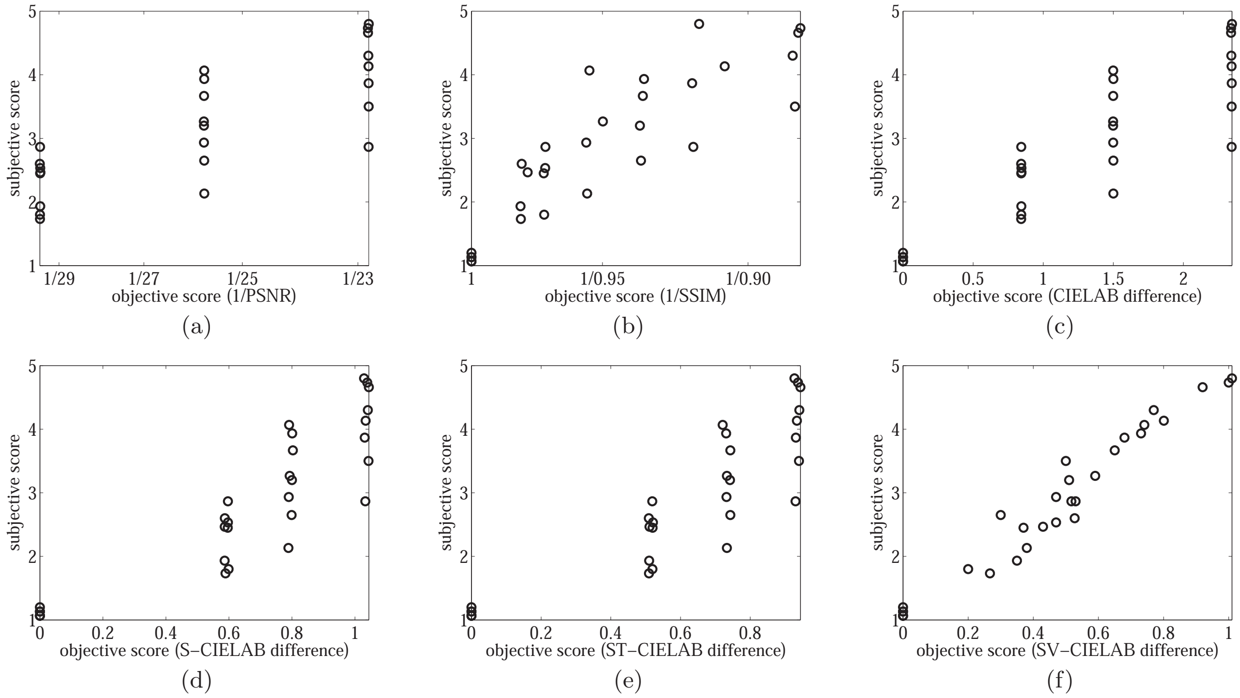


Fig. 13 Relationships between subjective scores and objective scores in experiment A. (a) PSNR. (b) SSIM. (c) CIELAB. (d) S-CIELAB. (e) ST-CIELAB. (f) SV-CIELAB. Note that the values of horizontal axes in (a) and (b) are reciprocals of PSNR and SSIM, respectively.

Table 2 Rank order correlation coefficients between subjective scores and objective scores in experiment A.

	MSE/PSNR/CIELAB	SSIM	S-CIELAB	ST-CIELAB	SV-CIELAB
correlations in Sample 1	0.783	0.869	0.811	0.837	0.915
correlations in Sample 2	0.817	0.889	0.853	0.871	0.938
correlations in all videos	0.792	0.878	0.833	0.861	0.925

4.2 Experiment B

In Experiment B, we prepared six kinds of videos whose velocities and directions of motion in each pixel were different. The motion of the videos is as described below, and Fig. 14 shows the example of the motion (Video#6).

Video#1: A man stand at center and a video camera moves around the man.
Video#2: A man stand and a video camera moves from left to right.
Video#3: The position of a video camera is fixed and a man walks from left to right.
Video#4: The position of a video camera is fixed and a man walks from far to near side.
Video#5: A man walks from left to right and a video camera moves with the motion of the man.
Video#6: A man walks from right to left and a video camera moves with the motion of the man.

In addition, we also prepared two more videos which have the same contents as the two videos (Video#5 and #6) in the above six videos, but the velocities of the added two videos were doubled. The velocities of the prepared videos were acquired by calculating the

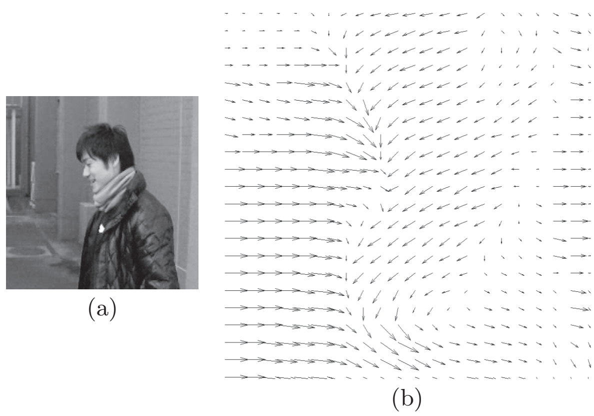


Fig. 14 Example of motion in a video prepared in Experiment B. (a) Frame extracted from a video sequence. (b) Estimated optical flows.

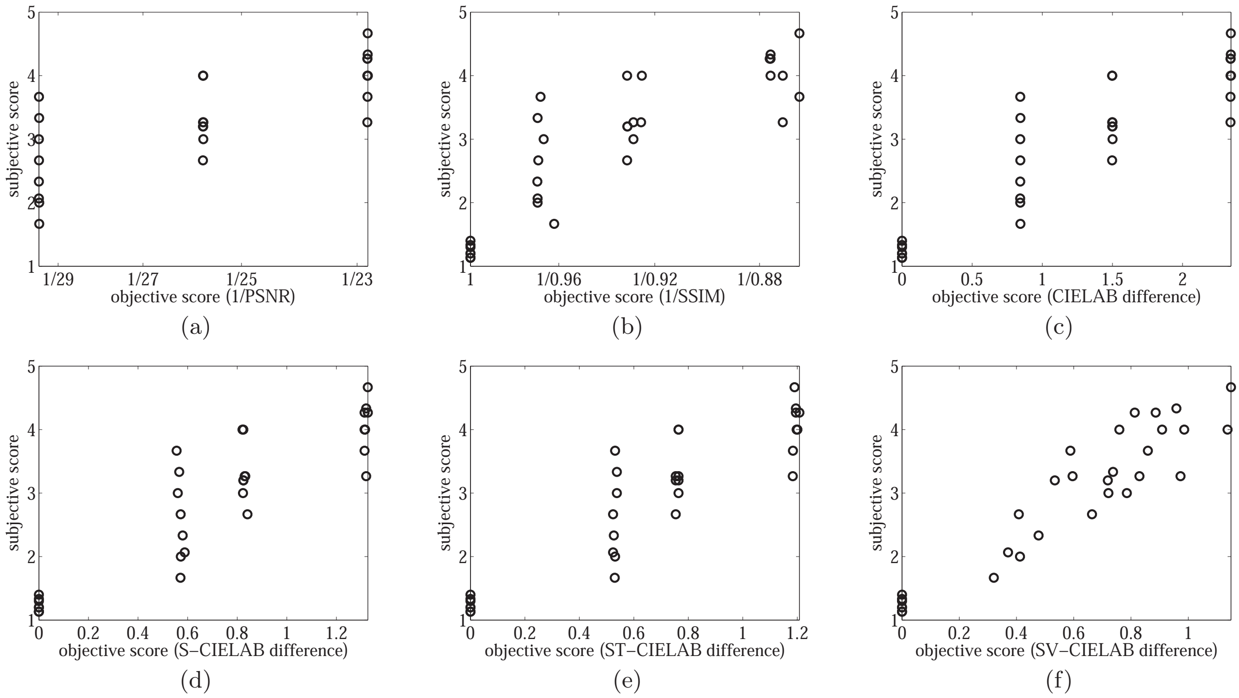


Fig. 15 Relationships between subjective scores and objective scores in experiment B. (a) PSNR. (b) SSIM. (c) CIELAB. (d) S-CIELAB. (e) ST-CIELAB. (f) SV-CIELAB. Note that the values of horizontal axes in (a) and (b) are reciprocals of PSNR and SSIM, respectively.

Table 3 Rank order correlation coefficients between subjective scores and objective scores in experiment B.

	MSE/PNSR/CIELAB	SSIM	S-CIELAB	ST-CIELAB	SV-CIELAB
correlations in all videos	0.771	0.866	0.821	0.843	0.906

optical flows as described in Section 3. The distorted videos were generated by adding random noise of 0, 7.5, 10 and 12.5% and totally 32 videos were evaluated. Other experimental set-ups are the same as the experiment A.

Figure 15 shows the results of the relationships between subjective and objective scores. Table 3 also shows the rank order correlation coefficients between subjective scores and the VQA methods. The characteristic features of these results are similar to those of the results in Experiment A, and SV-CIELAB is the more efficient VQA method than the conventional methods.

5. Conclusions

In this paper, we proposed SV-CIELAB which is a VQA method using a SV-CSF. From the experimental results for the validation, it was shown that the SV-CIELAB is a more efficient VQA method than the conventional methods which are PSNR, SSIM, CIELAB color difference, S-CIELAB and ST-CIELAB.

In SV-CIELAB, we assumed that observers' eyes tracked moving objects in videos while observers evaluate the video quality. However this assumption is not validated in this research. Therefore, as the future work, we should investigate the assumption by using the instrument to measure eye movements. As other future work, we should address the various types of degradation to validate SV-CIELAB, because degraded videos with random noise is only used in the experiments. We also would like to investigate the effects of errors of the optical flow extraction on SV-CIELAB, because the filter based on SV-CSF is significantly affected by the accuracy of the optical flow extraction. In addition, we should expand SV-CIELAB to color video quality assessment. In our method, mainly gray-scale videos are evaluated. However, to build more efficient VQA method, we have to address color videos. Furthermore, we would like to incorporate the visual attention model using motion information to SV-CIELAB. Information of visual attention is effective in image quality assessment [7, 8, 24]. In particular, motion information plays an important role in predicting visual attention [25]. Therefore, we should utilize visual attention areas based on motion information for a high-performance VQA method.

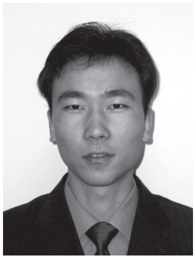
Acknowledgments

This work was partially supported by Grant-in-Aid for JSPS Fellows (21-5581) and Grant-in-Aid for Scientific Research (B) (19360026).

References

[1] Z. Wang and A.C. Bovik, Modern image quality assessment, Morgan & Claypool, 2006.
[2] B. Girod, "What's wrong with mean-squared error," in Digital Images and Human Vision, MIT Press, pp.207-220, 1993.
[3] T. N. Pappas, R. J. Safranek, and J. Chen, "Perceptual criteria for image quality evaluation," in Handbook of Image and Video Processing, Academic Press, pp.923-939, 2005.
[4] P. C. Teo and D. J. Heeger, "Perceptual image distortion," Proc. SPIE, Vol.2179, pp.127-141, 1994.
[5] S. Winkler, "A perceptual distortion metric for digital color video," Proc. SPIE, Vol.3644, pp.175-184, 1999.
[6] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. on Image Processing, Vol.13, No.4, pp.600-612, 2004.
[7] Z. Wang and X. L. Shang, "Spatial pooling strategies for perceptual image quality assessment," Proc. IEEE International Conference on Image Processing, pp.2945-2948, 2006.
[8] Z. Wang, A and Q. Li, "Video quality assessment using a statistical model of human visual speed perception," J. Opt. Soc. Am. A, Vol.24, Issue 12, pp.B61-B69, 2007.
[9] P. G. J. Barten, Contrast sensitivity of the human eye and its effects on image quality, SPIE Press, 1999.
[10] X. Zhang and B.A. Brainard, "A spatial extension of CIELAB for digital color image reproduction," SID Symposium Digest, Vol.27, pp.731-734, 1996.
[11] X.Tong, D. Heeger and C.B. Lambrecht, "Video quality evaluation using ST-CIELAB" , Proc. SPIE , Vol.3644, pp.185-196, 1999.
[12] D. H. Kelly, "Motion and vision. II. Stabilized spatio-temporal threshold surface," J. Opt. Soc. Am., Vol.69, Issue 10, pp.1340-1349, 1979.
[13] D. H. Kelly, "Spatiotemporal variation of chromatic and achromatic contrast thresholds," J. Opt. Soc. Am., Vol.73, Issue 6, pp.742-749, 1983.
[14] S. Yasukawa, T. Abe and H. Haneishi, "Quantification of color motion picture quality considering human visual sensitivity," Proc. IS&T's Fourth European Conference on Color in Graphics, Imaging and Vision, pp.116-119, 2008.
[15] S. Daly, "Engineering observations from spatiovelocity and spatiotemporal visual models," Proc. SPIE, Vol.3299, pp.180-191, 1998.
[16] J. Laird, M. Rosen, J. Pelz, E. Montag, and S. Daly, "Spatio-velocity CSF as a function of retinal velocity using unstabilized stimuli," Proc. SPIE, Vol.6057, 2006.
[17] K. Hirai, N. Tsumura, T. Nakaguchi and Y. Miyake, "Measurement and modeling of viewing-condition-dependent spatio-velocity contrast sensitivity function," Proc. IS&T and SID's 15th Color Imaging Conference, pp.106-111, 2007.
[18] J. R. Bergen, P. Anandan, K.J. Hanna and R. Hingorani, "Hierarchical model-based motion estimation," Proc. European Conference on Computer Vision, pp.237-252, 1992.
[19] J. Someya, and H. Sugiura, "Evaluation of liquid-crystal-display motion blur with moving-picture response time and human perception," J. Soc. Inf. Display, Vol.15, Issue 1, pp.79-86, 2007.
[20] K. Hirai, T. Nakaguchi, N. Tsumura and Y. Miyake, "Correlation analysis between motion blur width and human perception," Proc. 14th International Display Workshops, pp.1201-1204, 2007.
[21] ITU-R BT.500-11, Methodology for the subjective assessment of the quality of television pictures, 2002.
[22] C. Spearman, "The proof and measurement of association between two things," Amer. J. Psychol., Vol.15, pp.72-101, 1904.
[23] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," http://www.vqeg.org/, 2000.
[24] J. Bai, T. Nakaguchi, N. Tsumura, and Yoichi Miyake, "Evaluation of image corrected by retinex method based on S-CIELAB and gazing information," IEICE Trans. on Fundamentals, Vol.E89-A, No.11, pp.2955-2961, 2006.
[25] J. M. Wolf, and T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?," Nature Reviews Neuroscience, Vol.5, No.6, pp.1-7, 2004.

Keita Hirai received the B.E. and M.S. degrees from Chiba University in 2005 and 2007. He is currently a Ph.D course student in Chiba University. He is also curentlty a research fellow of Japan Society for the Promotion of Science (JSPS). He is interested in researches for evaluation and improvement of video quality using human visual characteristics. He is a student member of the Institute of Electronics, Information and Communication Engineers (IEICE), the Institute of Image Information and Television Engineers (ITE), Japan and Society of Photographic Science and Technology of Japan (SPSTJ), ACM SIGGRAPH, IS&T, SID and SPIE.



Jambal Tumurtogoo received the B.E. degree from Chiba University in 2009. He is currently a master's course student in Chiba University. He is a student member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Ayano Kikuchi received the B.E. degree from Chiba University in 2008. She is currently a master's course student in Chiba University. Her research interests include room illumination while viewing displays.

Norimichi Tsumura received the B.E., M.E. and Dr. Eng degrees in Applied Physics from Osaka University in 1990, 1992 and 1995, respectively. He moved to the Department of Information and Computer Sciences, Chiba University in April 1995, as an Assistant Professor. He is currently Associate Professor in Department of Information and Image Sciences, Chiba University since February 2002. He got the Optics Prize for Young Scientists (The Optical Society of Japan) in 1995, Applied Optics Prize for the excellent research and presentation (The Japan Society of Applied Optics) in 2000, Charles E. Ives Award (Journal Award: IS&T) in 2002, 2005. He is interested in the color image processing, computer vision, computer graphics and biomedical optics.

Toshiya Nakaguchi received the B.E., M.E., and Ph.D. degrees from Sophia University,Tokyo, Japan in 1998, 2000, and 2003, respectively. He was a research fellow supported by Japan Society for the Promotion of Science from April 2001 to March 2003. From 2006 to 2007, he was a research fellow in Center of Excellence in Visceral Biomechanics and Pain, in Aalborg Denmark, supported by CIRIUS, Danish Ministry of Education from 2006 to 2007. Currently, he is an Assistant Professor of imaging science at the Graduate School of AdvancedIntegration Science, Chiba University, Chiba, Japan. His current research interests include the computer assisted surgery and medical training, medical image analysis, real-time image processing, and image quality evaluation. He is a member of the IEEE, IS&T, the Institute of Electronics, Information and Communication Engineers (IEICE), Japan and Society of Photographic Science and Technology of Japan.

Yoichi Miyake has been professor of Chiba University since 1989. He received Ph.D. from Tokyo Institute of Technology in 1978. During 1978 and 1979, he was a post doctoral fellow of Swiss Federal Institute of Technology (ETHZ). In 1997, he was a guest professor of University of Rochester. He received Charles E Ives Award (paper award) from IS&T in 1991, 2001 and 2005. He became a fellow of IS&T in 1995. He was named as Electronic Imaging Honoree of the year in 2000 from SPIE and IS&T. He became honorary member of IS&T in 2003. He published many books and original papers on the image processing, color science, image evaluations and he is known as a pioneer of spectral image processing. He was served as a president of SPSTJ (The Society of Photographic Science and Technology of Japan) from 2000 to 2002 and a vice president of IS&T (The Society for Imaging Science and Technology, USA) from 2000 to 2004. He was also served as a president of The Japanese Association of For ensic Science and Technology from 1998 to 1999. From 2003 to 2009, he was served as professor and director of Research Center for Frontier Medical Engineering in Chiba University. He is currently served as research professor in Chiba University.