SV-CIELAB: Video Quality Assessment using Spatio-Velocity Contrast Sensitivity Function

Keita Hirai [1], Jambal Tumurtogoo [1], Ayano Kikuchi [1], Toshiya Nakaguchi [1], Norimichi Tsumura [1] and Yoichi Miyake [2]
1. Graduate School of Advanced Integration Science, Chiba University, Chiba, Japan
2. Research Center for Frontier Medical Engineering, Chiba University, Chiba, Japan

Abstract

In this paper, we proposed and validated SV-CIELAB which is a video quality assessment (VQA) method using a spatio-velocity contrast sensitivity function (SV-CSF). The SV-CSF consists of the relationship among contrast sensitivities, spatial frequencies and velocities of stimuli. We used the SV-CSF for filtering original and distorted videos. The criteria in our method are obtained by calculating CIELAB color differences between filtered videos. From the experimental results for the validation, it was shown that SV-CIELAB is the more efficient VQA method than conventional methods such as CIELAB color difference, Spatial-CIELAB and so on.

Introduction

Recently, according 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.

Since human are the ultimate observers for evaluating 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 hands, an objective VQA method is not expensive compared with a subjective method. In general, factors of image quality in objective evaluations are quantified by the criteria such as sharpness (resolution), color reproduction, tone reproduction and noise characteristics. Moreover, since images are perceived through the human visual system, objective VQA methods should be designed by incorporating human visual characteristics [1].

In conventional image quality assessment methods with the human visual system, contrast sensitivity functions (CSF) are frequently used. In 1996, Zhang et al. proposed Spatial-CIELAB (S-CIELAB) [2]. To compute image degradations between original and distorted images, they applied a CSF for filtering images. The criteria can be obtained by calculating CIELAB color differences between filtered original and distorted images. Though S-CIELAB is useful for still image quality assessment, temporal characteristics in human visual system were not considered and it is not appropriate for VQA. To evaluate video quality, Tong et al, proposed spatio-temporal CIELAB (ST-CIELAB) [3] which is an extension of S-CIELAB. They used a spatio-temporal CSF [4][5] for filtering videos. Though ST-CIELAB was proposed as a VQA method, a previous research has reported that ST-CIELAB is not more efficient than S-CIELAB [6]. 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.

Therefore, in this paper, we propose SV-CIELAB which is a VQA method using a spatio-velocity contrast sensitivity function (SV-CSF) [5][7]. The SV-CSF consists of the relationship among contrast sensitivities, spatial frequencies and velocities of stimuli. It also contains eye movement characteristics. We used the SV-CSF for filtering original and distorted videos. The criteria in our method are obtained by calculating CIELAB color differences between filtered videos. Furthermore we performed subjective experiments for validating SV-CIELAB. The subjective results were compared with SV-CIELAB and the conventional VQA methods which are PSNR, SSIM [8], CIELAB color difference, S-CIELAB and ST-CIELAB.

Related Works

Spatial-CIELAB and Spatio-Temporal CIELAB

In S-CIELAB or ST-CIELAB [2][3], 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.

Spatio-Velocity Contrast Sensitivity Function

Figure 1 represents the SV-CSF model [7]. The SV-CSF consists of the relationship among contrast sensitivities, spatial frequencies and velocities of stimuli. For modeling the SV-CSF, the visual lines of the observers in the experiments followed the moving stimuli and contrast sensitivities were measured. 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.

SV-CIELAB

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

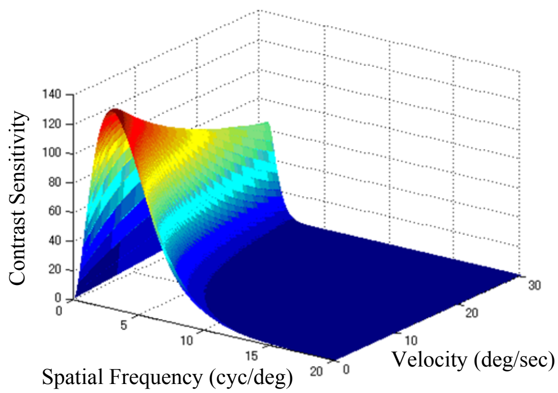


Figure 1 *SV-CSF model.*

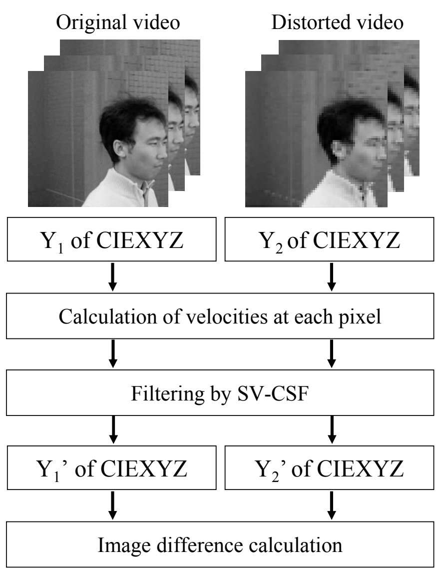


Figure 2. *Overview of SV-CIELAB.*

color space are used in the processing of the videos. Figure 2 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 [9]. 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.

Filtering in SV-CIELAB

Figure 3 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. 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 obtain video frames separated in spatial frequency domain. Each separated frame has the information of one cyc/deg. By using

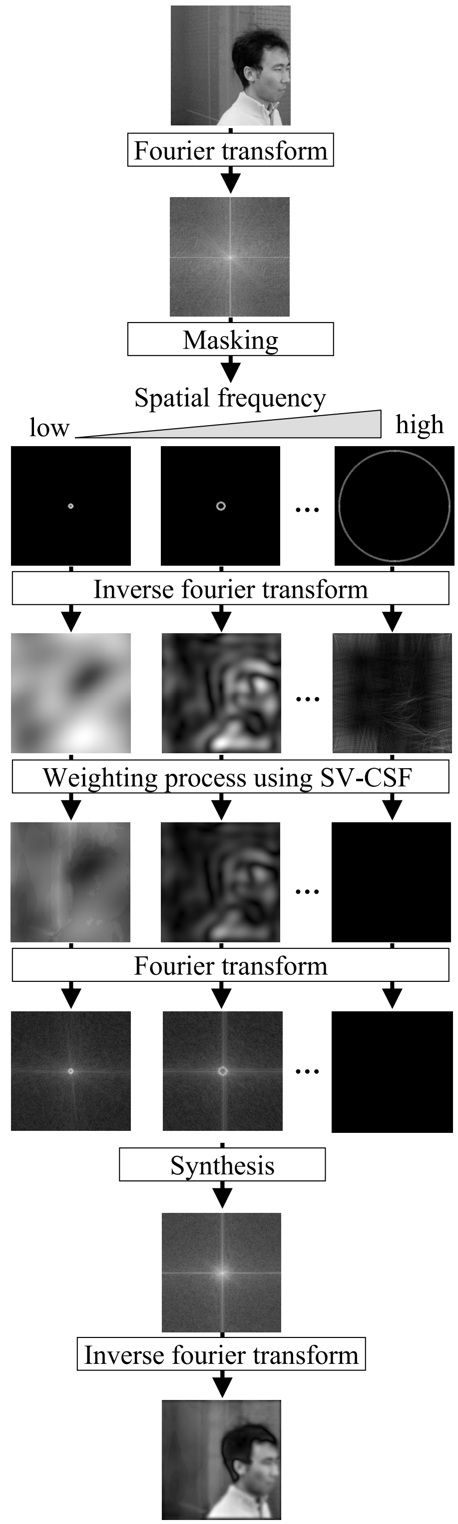


Figure 3. *Filtering process in SV-CIELAB.*

velocity information, the frames separated by each spatial frequency are weighted by contrast sensitivities in the SV-CSF model (described in the next section). A final filtered frame is obtained by synthesizing the weighted frames.

©2009 Society for Imaging Science and Technology

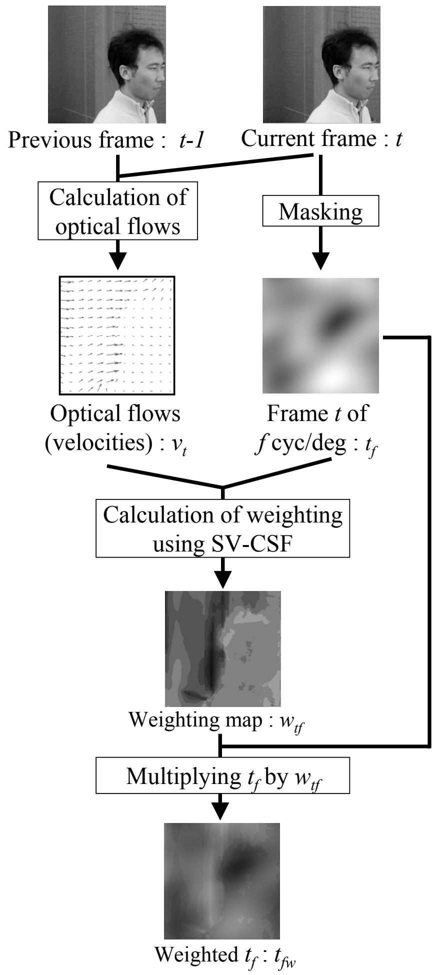


Figure 4. *Weighting process using SV-CSF.*

Weighting Map

Figure 4 shows the weighting process using the SV-CSF. In this process, weighing map *w* at frame *t* configured by specific spatial frequency *f* (cyc/deg) is generated by a following equation.

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

where *x* and *y* are spatial coordinates, v_t is velocity (deg/sec) at frame *t*. $\text{SV-CSF}_{\text{nor}}$ 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 the previous section. The SV-CSF [7] is computed by

$$\begin{aligned} &\text{SV-CSF}(f,v) = k \cdot c_0 \cdot c_1 \cdot c_2 \cdot (v + c_v) \cdot (c_1 \cdot 2\pi\rho)^2 \exp\left(-\frac{c_1 \cdot 4\pi f}{\rho_{\max}}\right) \\ &k = 6.1 + 7.3 \cdot \left|\log(c_2(v + c_v)/3)\right|^3 \\ &\rho_{\max} = 45.9/(c_2(v + c_v) + 2) \end{aligned} \tag{2}$$

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

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

(a)Sample 1: forest *(b)Sample 2: standing men*

Figure 5. *Displayed videos.*

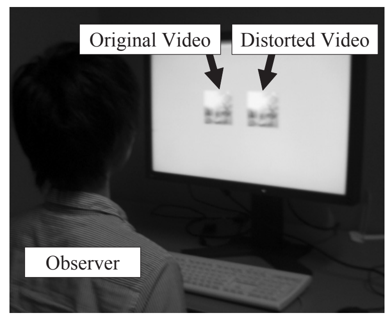


Figure 6. *Experimental room.*

Image Difference Calculation

The criteria in SV-CIELAB are obtained by calculating CIELAB color differences between filtered original and distorted videos.

Validation

For the validation of SV-CIELAB, two kinds of subjective evaluation experiments were performed. In this validation, subjective experimental results were compared with SV-CIELAB and the conventional VQA methods which are PSNR, SSIM [8], CIELAB color difference, S-CIELAB and ST-CIELAB.

Experiment A

In Experiment A, we prepared two videos (Sample 1 and 2 shown in Fig. 5) which velocities and directions of motion in each pixel 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%. 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. Actually, in the experiments, the video of 20 seconds were presented by running a video of 10 seconds continuously.

Figure 6 shows the experimental room. Fifteen observers participated in Experiment A. The original and distorted videos were displayed at the same time, and the observers evaluated the degraded level by the scale of 1 to 5 (1: the same ~ 5: very different). The display device is a 27” LCD (1920 × 1200 pixels) and the viewing distance is 400mm which are decided based on the standard viewing distance recommended by ITU [10].

Figure 7 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 and they evaluate approximately slightly different or different (the subjective scores are 2 or 3) between the original and distorted videos.

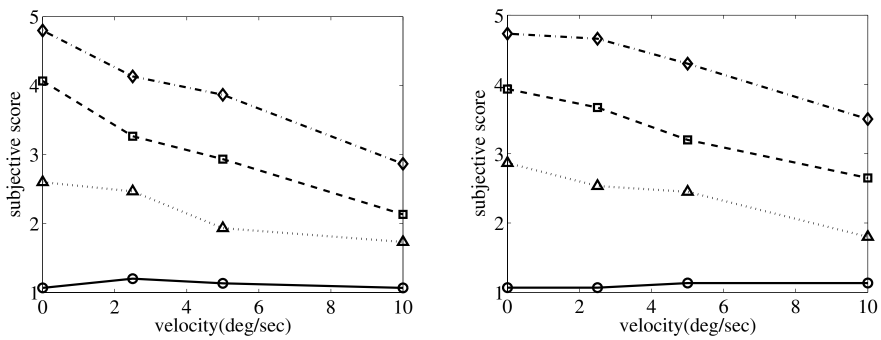
Figure 8 and 9 show the results that describe the relationships between the objective scores and the velocities in the videos (Sample 1 and 2). As shown in Fig. 8 and 9, the objective scores of the conventional VQA methods are approximately the same values in each noise level respectively. These results mean that the conventional methods practically are not affected by the changes of the velocities. On the other hand, the objective scores in SV-CIELAB become lower values as the velocities increase. The tendency of SV-CIELAB is similar to the subjective results shown in Fig. 7.

Figure 10 shows the results which represents the relationships between subjective and objective scores. Table 1 also shows Spearman's rank order correlation coefficients (ROCCs) [11] between subjective scores and the proposed and conventional VQA methods. As shown in Fig. 10 and Table 1, SV-CIELAB is the more efficient VQA method than the conventional methods.

Experiment B

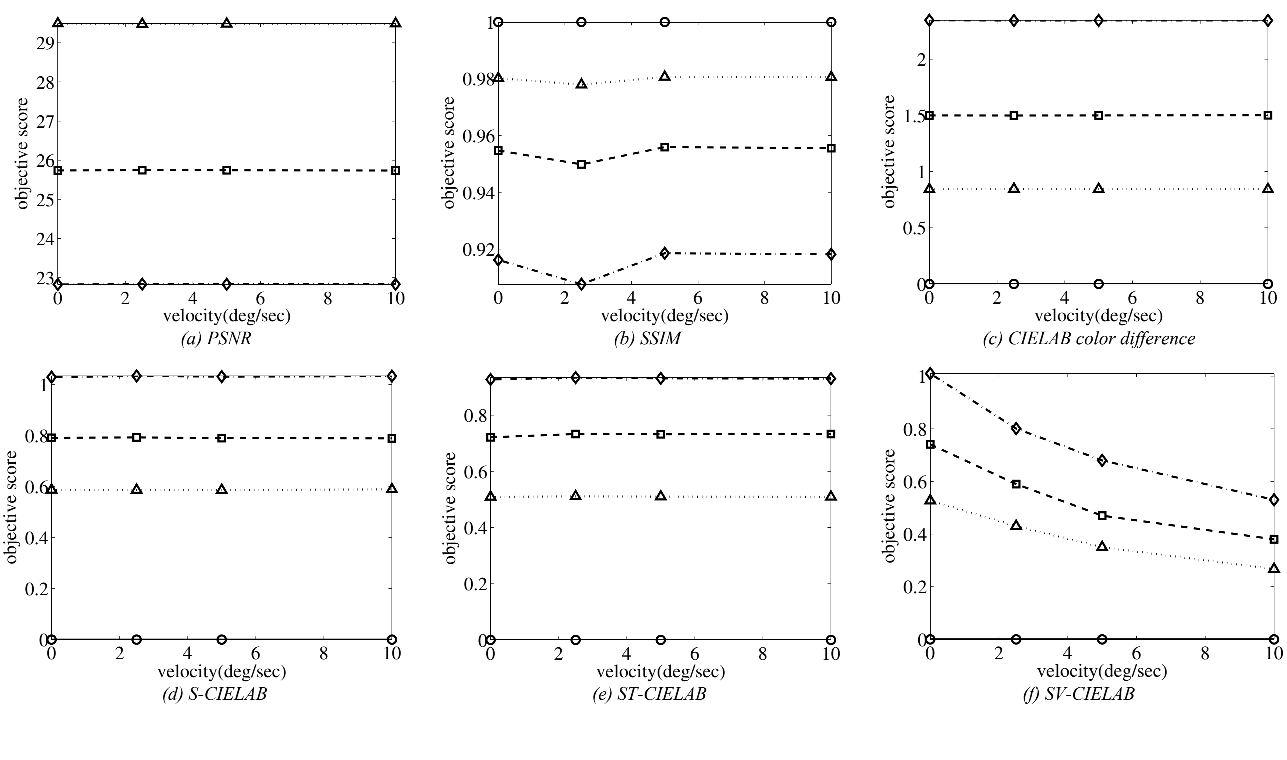
In Experiment B, we prepared eight kinds of typical videos which velocities and directions of motion in each pixel were different. The velocities were acquired by calculating the optical flows as described before. The distorted videos were generated by adding random noise of 0, 7.5, 10 and 12.5% and totally 32 videos were evaluated. Other experimental setups are the same as Experiment A.

Figure 11 shows the results of the relationships between subjective and objective scores. Table 2 also shows ROCCs between subjective scores and the VQA methods. From the results shown in Fig. 12 and Table 2, SV-CIELAB is the more efficient VQA method than the conventional methods. These results are similar to the results of Experiment A.



(a) Sample 1 *(b) Sample 2*

Figure 7. *Relationships between subjective scores and velocities of videos. Noise levels are ○:0%, △:7.5%, □:10% and ◇:12.5%.*

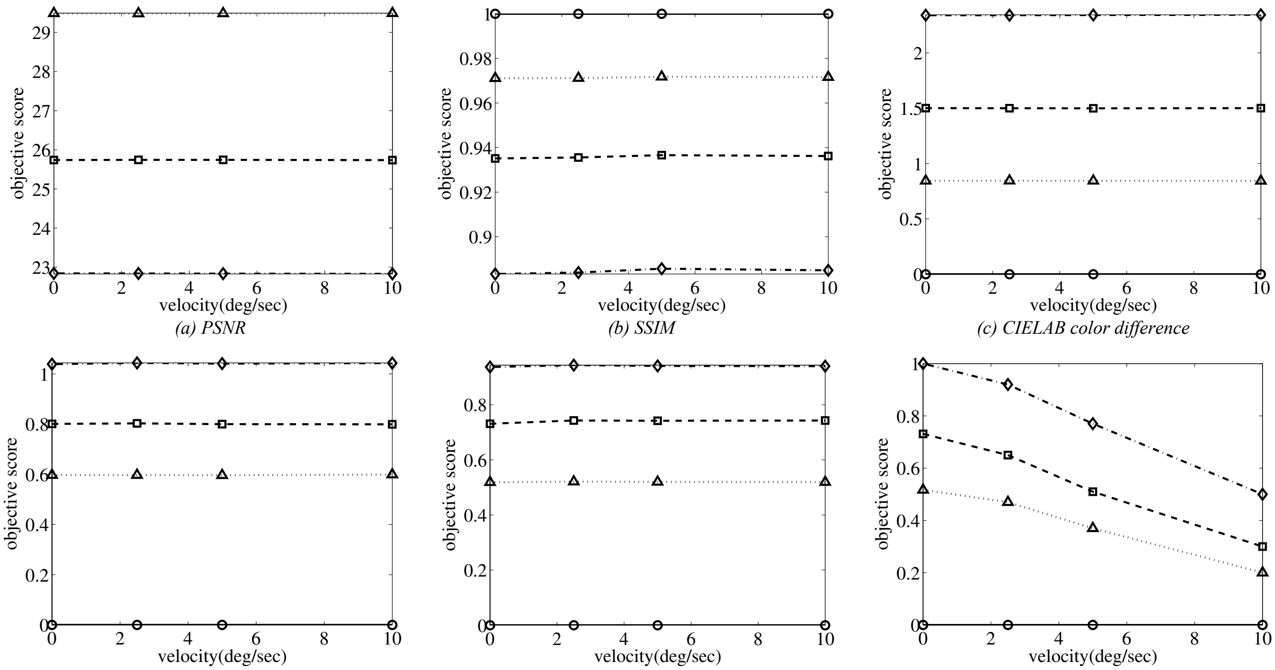


(a) PSNR *(b) SSIM* *(c) CIELAB color difference*

(d) S-CIELAB *(e) ST-CIELAB* *(f) SV-CIELAB*

©2009 Society for Imaging Science and Technology

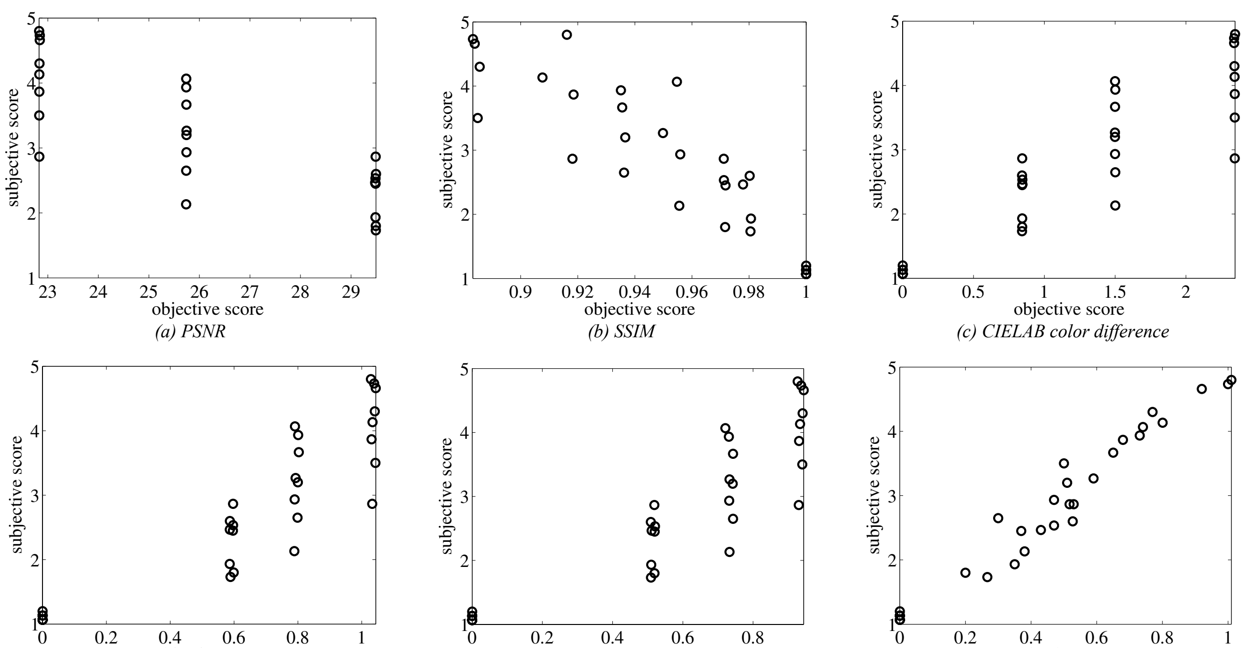
***Figure 8.** Relationships between objective scores and velocities in Sample 1. Noise levels are ○:0%, △:7.5%, □:10% and ◇:12.5%.*



(a) PSNR *(b) SSIM* *(c) CIELAB color difference*

(d) S-CIELAB *(e) ST-CIELAB* *(f) SV-CIELAB*

***Figure 9.** 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 color difference*

(d) S-CIELAB *(e) ST-CIELAB* *(f) SV-CIELAB*

***Figure 10.** Relationships between subjective scores and objective scores in Experiment A.*

Table 1. Rank order correlation coefficients between subjective and objective scores in Experiment A

	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

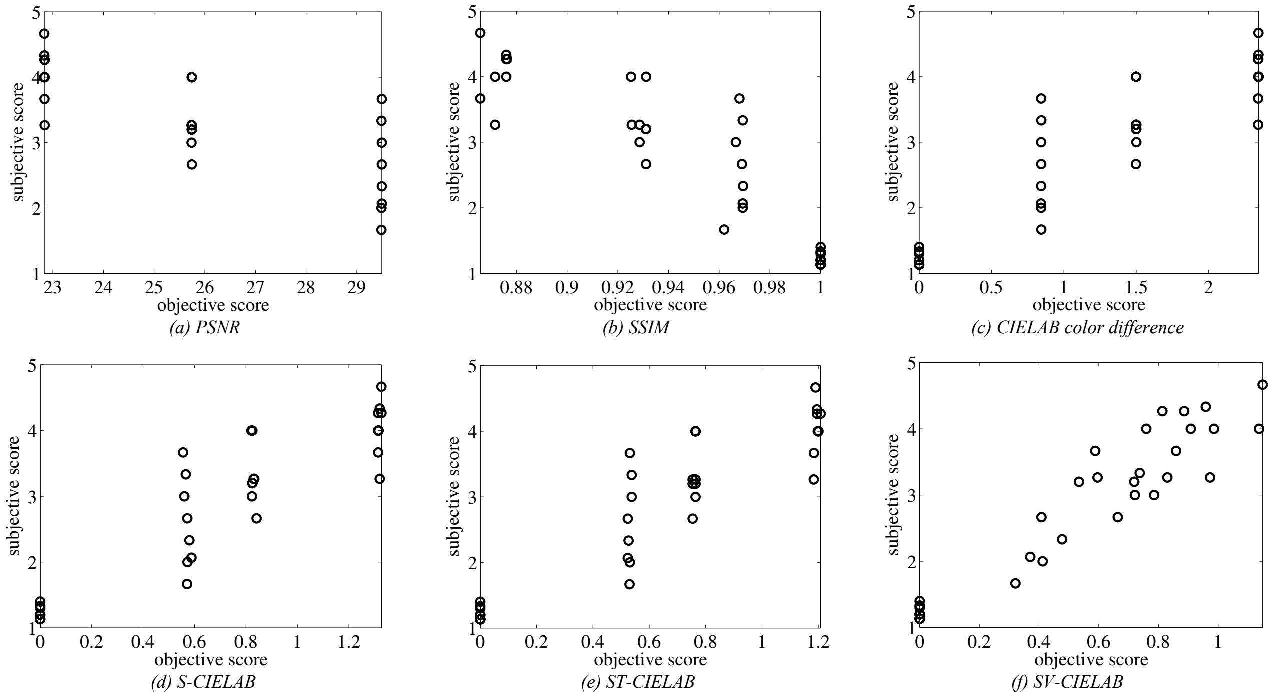


Figure 11. *Relationships between subjective scores and objective scores in Experiment B.*

Table 2. Rank order correlation coefficients between subjective and objective scores in Experiment B

	PSNR / CIELAB	SSIM	S-CIELAB	ST-CIELAB	SV-CIELAB
Correlations in all videos	0.7711	0.866	0.821	0.843	0.906

Conclusion

In this paper, we proposed SV-CIELAB which is a video quality assessment method using a spatio-velocity contrast sensitivity function. 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.

As the 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. In addition, we would like to 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.

References

[1] Z. Wang and A.C. Bovik, Modern Image Quality Assessment• Morgan & Claypool (2006).

[2] 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).

[3] X.Tong•D. Heeger and C.B. Lambrecht "Video Quality Evaluation using ST-CIELAB", Proc. SPIE •Vol.3644, pp.185-196 (1999).

[4] D.H. Kelly, Motion and vision. II. Stabilized spatio-temporal threshold surface, J. of Optical Society of America, Vol.69, pp.1340-1349 (1979).

[5] S. Daly, "Engineering observations from spatiovelocity and spatiotemporal visual models," Proc. SPIE, Vol.3299, pp.180-191 (1998).

[6] S. Yasukawa, T. Abe and H. Haneishi, "Quantification of color motion picture quality considering human visual sensitivity," IS&T's Fourth European Conference on Color in Graphics, Imaging and Vision, pp.116-119 (2008).

[7] K. Hirai, N. Tsumura, T. Nakaguchi and Y. Miyake, "Measurement and Modeling of Viewing-Condition-Dependent Spatio-Velocity Contrast Sensitivity Function," IS&T and SID's 15th Color Imaging Conference, pp.106-111 (2007).

[8] 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).

[9] J.R. Bergen, P. Anandan, K.J. Hanna and R. Hingorani, "Hierarchical model-based motion estimation," Proc. of *the European Conference on Computer Vision*, pp. 237–252 (1992).

[10] ITU-R BT.500-11, Methodology for the subjective assessment of the quality of television pictures, 2002.

[11] C. Spearman, “The proof and measurement of association between two things,” Amer. J. Psychol., Vol.15, pp.72–101 (1904).

Author Biography

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 currently 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.