A psychophysical study of improvements in motion-image quality by using high frame rates

Y. Kuroki
T. Nishi
S. Kobayashi
H. Oyaizu
S. Yoshimura

Abstract — The ideal frame rate for the highest motion-image quality with respect to blur and jerkiness is presented. In order to determine the requirements for avoiding these impairments, motion images from a high-speed camera and computer graphics were combined with a high-speed display to perform a psychophysical evaluation. The camera, operating at 1000 fps, and image processing were used to simulate various frame rates and shutter speeds, and a 480-Hz CRT display was used to present motion images simulating various frame rates and time characteristics of the display. Subjects were asked to evaluate the difference in quality between motion images at various frame rates. A frame rate of 480 fps was chosen to be an appropriate reference frame rate that, as a first estimation, enables coverage up to the human-dynamic-resolution (HDR) limit based on another experiment using real moving charts. The results show that a frame rate of 120 fps provides good improvement compared to that of 60 fps, and that the maximum improvement beyond which evaluation is saturated is found at about 240 fps for representative standard-resolution natural images.

Keywords — *Vision science, human factors, frame rate, motion-image quality.*

1 Introduction

How precisely can a moving object be seen and how fast can that object be moving? And how about a reproduced motion image from a real moving object? We launched our investigation related to motion-image quality based on such basic questions. In addition to studying the resolution limit of smooth pursuit for a motion image, we also evaluated the perception of blur and jerkiness.

To answer the first question, we performed psychophysical experiments[1] to determine the upper limit of the lateral velocity of real visual-target charts at which they can be resolved with smooth pursuit. Westheimer (1954)[2] reported that the human eye can follow a visual target with the same velocity as the target velocity if it is less than 30 dps (degrees per second). Research later found that the smooth-pursuit velocity is less than the visual-target velocity at various speeds. Meyer *et al.* (1985),[3] reported that the ratio of the pursuit velocity to the target velocity is 0.87, and that the upper limit of smooth pursuit is 100 dps, from their experiments with trained subjects. They used a beam spot (not spatial-frequency charts) as the visual target. In our measurements, the upper limit of the velocity of the visual target at which it can be resolved with smooth pursuit was determined for targets in various spatial frequencies.

To answer the second question, we used a high-speed camera and computer graphics together with a high-speed display to present reproduced motion images. Again, we performed psychophysical experiments by asking subjects to evaluate the difference in quality between motion images at various frame rates. They were asked to evaluate the motion image quality in terms of blur and jerkiness. Jerkiness is defined in an ANSI standard[4] as "motion that was originally smooth and continuous and is perceived as a series of distinct snapshots." We utilized a high-speed camera (nac Imaging Technology, MEMRECAM) operating at 1000 fps. Both blur and jerkiness are expected to depend on both the frame rate and shutter speed of the camera and also on the frame rate and duty ratio of the display. In our study, image processing was used to simulate various frame rates of the camera and display from 62 to 500 fps and camera shutter speeds which either correspond to an open shutter at each frame rate or is fixed at 1/500 sec (by frame averaging). A 480-Hz CRT display was used to present the processed motion images in a "hold-type" manner by a repeated display of the same 1/480-sec CRT frame rate during the simulated 62–500-fps frame interval.

2 Resolution limit of smooth pursuit for a motion image

For our first experiment, to study the resolution limit of smooth pursuit using spatial-frequency charts, we compared the case of real charts (directly looking-at-the-screen measurement) and for film and video reproductions (looking-at-the-film and video-projected-image measurements).

We first constructed a "Screen mover," the large moving screen shown in Fig. 1. The screen size was 1 m high and 6 m wide. The screen was a white polyester film and was moved horizontally by servo (Servoland SVMM300, SVEM5-A) controlled rollers.

In the directly looking-at-the-screen measurement, we masked around a window of a 20° horizontal viewing

Revised extended version of a paper presented at the 2006 SID Symposium held June 6–9, 2006 in San Francisco, California.

The authors are with Sony Corp., Technology Development Group, 1-11-1 Osaki, Shinagawa-ku, Tokyo 141-0032, Japan;
telephone +81-3-5435-3302, fax -3380, e-mail: Yoshihko.Kuroki@jp.sony.com.

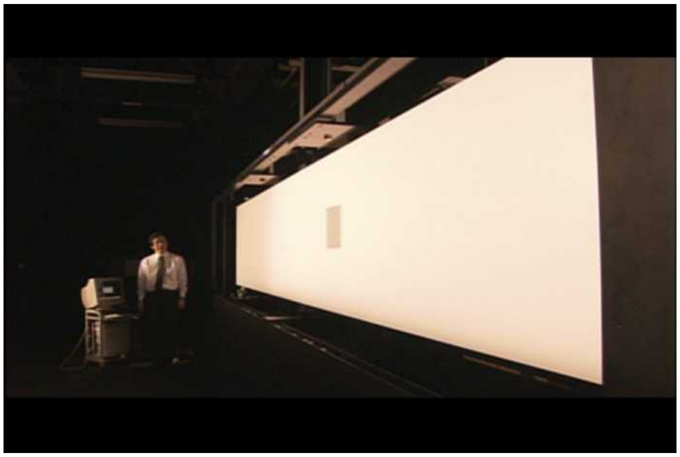


FIGURE 1 — Visual stimulus moving equipment: "Screen mover."

FIGURE 3 — Screen mover and screen for projection.

angle (1.728 m) and 3.82° for a vertical viewing angle (0.333 m) at a viewing distance of 5 m with black panels due to the maximum speed of the Screen mover. The maximum luminance of the screen was 40 cd/m^2 using a metal-halide light (ARRI HMI 2.5 kW) in the non-flicker mode.

For the looking-at-the-film and video-projected-image measurement, we masked around a window for a 20° horizontal viewing angle (5.284 m) and 3.82° for a vertical viewing angle (1 m) at a viewing distance of 15 m using black panels, and we shot the moving screen using the spatial-frequency chart (which will be described later) using a 35-mm-film camera (ARRIFLEX 435 with Carl Zeis Vista 25 mm) with 180° and 45° shutters; 24p HD camera (Sony HDW-F900 with Canon HJ18 × 7.8) with 1/96″ and 1/198″ shutter and 60i HD camera (Sony HDW-700A with Canon HJ15 × 8) with also 1/96″ and 1/198″ shutters in appropriate distance and light (ARRI HMI 2.5 kW) conditions. And we projected the film and HD video images using a film projector and a digital light processing (DLP) projector (Sony LE-100), respectively, on a screen which was placed next to the screen mover. The projected image size of both film and HD video were the same size as the Screen mover. Both the maximum luminance of the film and HD video images was 40 cd/m^2. The experimental room for which both of the measurements above were performed in had a black wall and floor and no light condition (except for the projector).

With regard to the difference in viewing distance for the two types of measurements, we tested "static" resolution limits for five adults at 5-m and 15-m distances using the same frequency charts which we used in our dynamic measurements and confirmed that there is no significant difference.

The frequency charts we used were vertical black and white square wave stripes with a 50% duty ratio, and the size of the chart was 1° high and 1° wide (0.0873-m square for 5-m viewing distance and 0.262-m square for 15-m viewing distance). The variations in the spatial frequency of the charts were 150.6, 121.2, 94.8, 75.3, 59.7, 47.4, 37.8, 30.0, 23.7, 18.9, 15.0, 12.0, 9.3, and 7.5 cpd (cycles per second) for both sizes of the chart. Samples of the charts (both sizes at 12.0 cpd) are shown in Fig. 2.

In the directly looking-at-the-screen measurement, the frequency chart was attached to the screen and moved horizontally at one of the speed settings in 5-dps steps between 0 and 80 dps. The white panel, made of the same material as the screen to keep the visual adaptation, was placed over the window in the black panel covering the Screen mover. The white cover panel was opened to expose

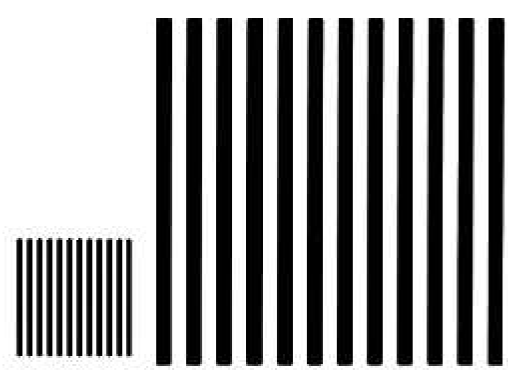


FIGURE 2 — Sample of frequency chart (12 cpd): left for 5 m, right for 10 m.

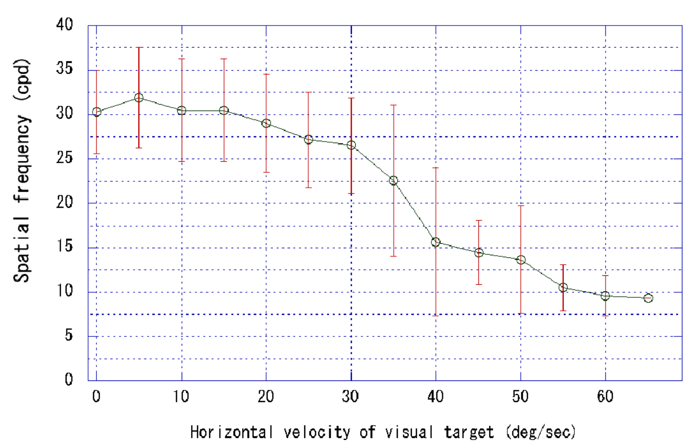


FIGURE 4 — Resolution limit in smooth pursuit for the moving actual (printed) visual target.

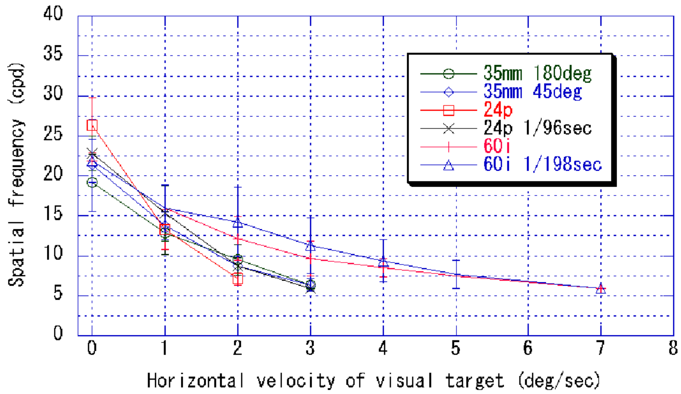


FIGURE 5 — Resolution limit in smooth pursuit for the moving projected visual target.

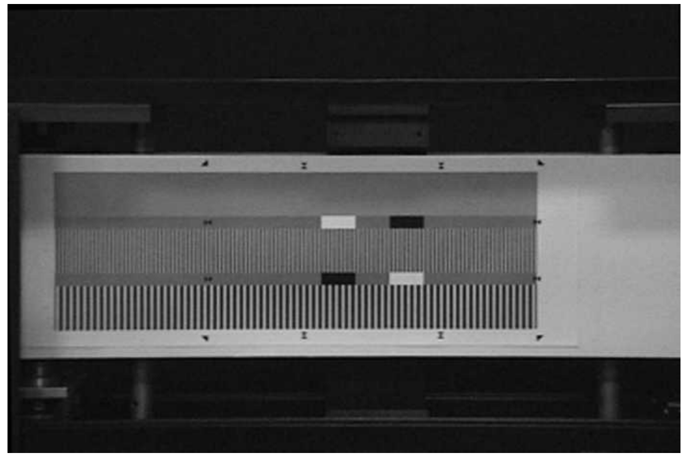


FIGURE 7 — Small type of screen mover and sinusoidal chart.

the moving chart to the subject three times, and the subject was asked if he could see the lines or just gray. The subjects were allowed to move their heads freely in order to see the moving screen. We tested by increasing the speed from low to high speed and by decreasing from high to low speed alternatively. When we found the near-threshold range, the speed sequence was changed to a random sequence. Each frequency chart at each speed was shown to the subject more than twice in one run. Each of the five subjects performed two runs on different days. We defined the maximum spatial frequency of the chart which subjects could resolve to be the resolution limit. We then made equivalent measurements of reproduced motion images using both the film and the HD video reproduction systems with projection equipment. A screen for the same-size projection as the "Screen mover" is shown in Fig. 3. The subjects were also five adults.

The human-resolution limit curve in Fig. 4 was obtained by looking directly at the chart on the screen using the psychophysical method. From 20°/sec, the resolution limit decreases gradually and at a low frequency such as 7.5 cycles per degree, the resolution limit reaches almost 60°/sec. Figure 5 shows the results of the film (35 mm) projection and the video (DLP) projection images. When the chart is motionless, the spatial frequencies of these resolution limits are not that different. However, the spatial frequencies of the resolution limit of the film and the video-projection images quickly decrease when the chart moves, and finally these limits are 3–7°/sec, corresponding to camera frame rates of 24 and 60 fps, respectively. The difference between looking-directly-at-the-screen and looking-at-the-film and video-projection images is more than 8 times (Fig. 6).

In order to separate the effect of retinal slip as a cause of blur, the contrast ratio of the 60i HD video output signal was measured by a waveform monitor for the horizontally moving sinusoidal chart (Fig. 7). Figure 8 shows that a higher velocity of the object or a lower shutter speed of the camera results in lower resolution, although subjective evaluation does not show such dependency on shutter speed and resolution in Fig. 5. This result shows that retinal slip in hold-type displays is a dominant factor as a cause of blur.

3 Perception of blur and jerkiness

In our next experiment, we evaluated blur and jerkiness for various frame rates using the psychophysical method. Figures 9–15 are frames from the stimulus motion images

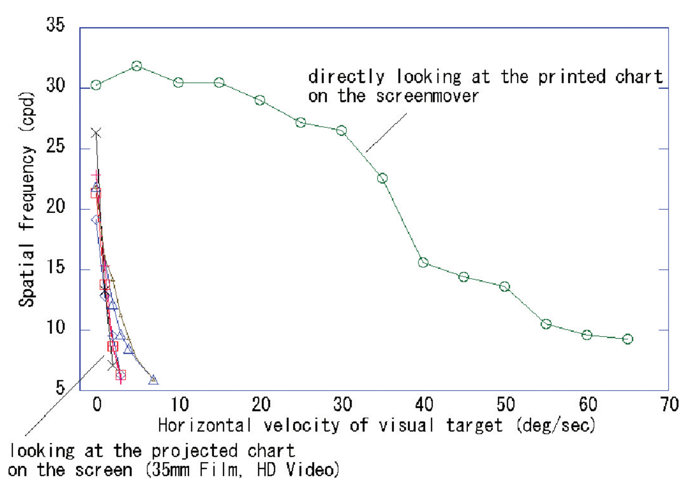


FIGURE 6 — Plot comparing the results shown in Figs. 4 and 5.

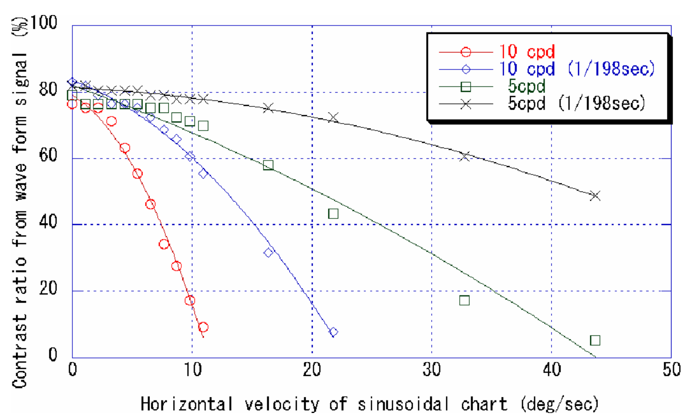


FIGURE 8 — Contrast ratio of HD video output (60i) on horizontal velocity of the sinusoidal chart.

FIGURE 9 — Ball.

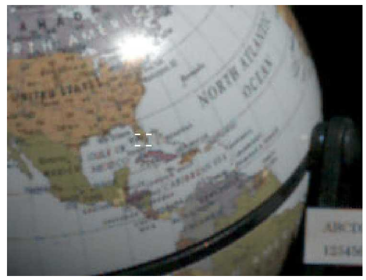


FIGURE 10 — Earth.

FIGURE 11 — Car.

FIGURE 12 — Garden.

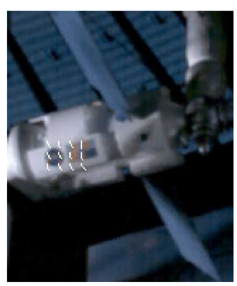


FIGURE 13 — Space.

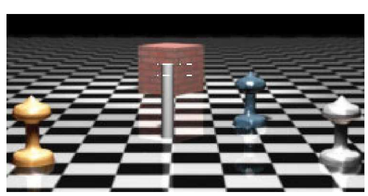


FIGURE 14 — Computer graphics.

FIGURE 15 — Screen mover.

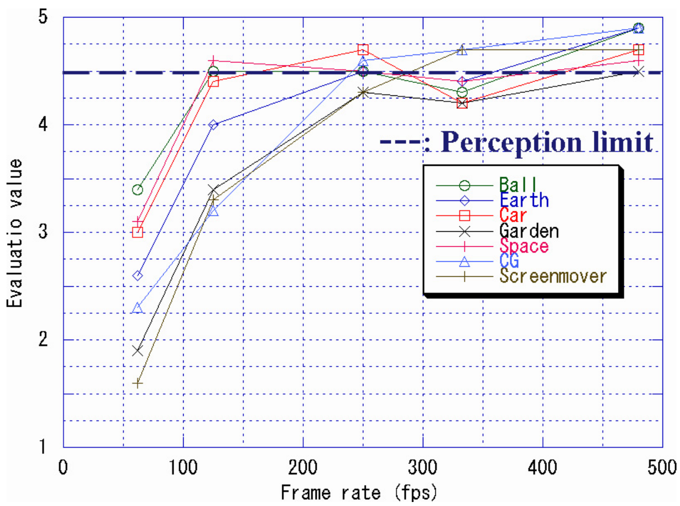


FIGURE 16 — Evaluation of motion-image quality for blur in pursuit viewing with open-shutter images.

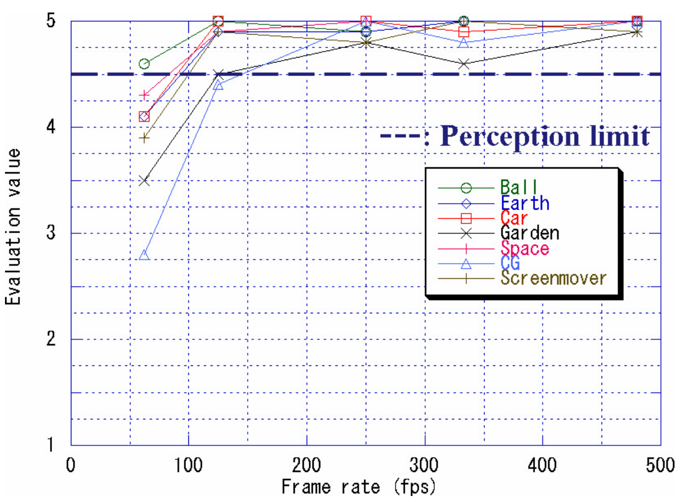


FIGURE 17 — Evaluation of motion-image quality for jerkiness in fixation viewing with open-shutter images.

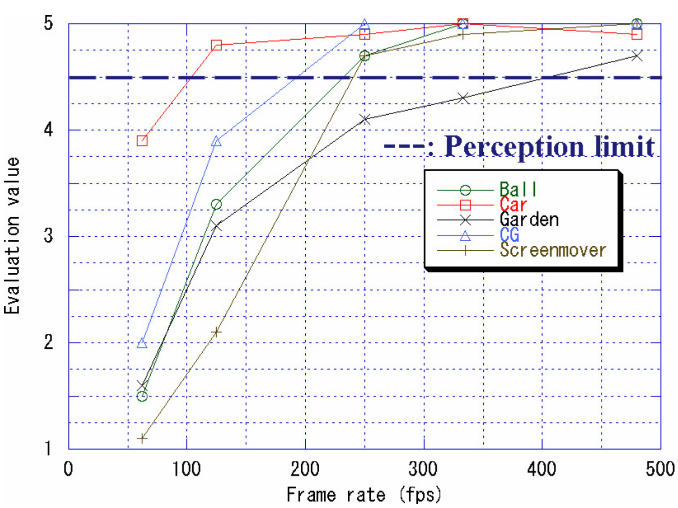


FIGURE 18 — Evaluation on motion-image quality for jerkiness in fixation viewing with 1/500-sec shutter images.

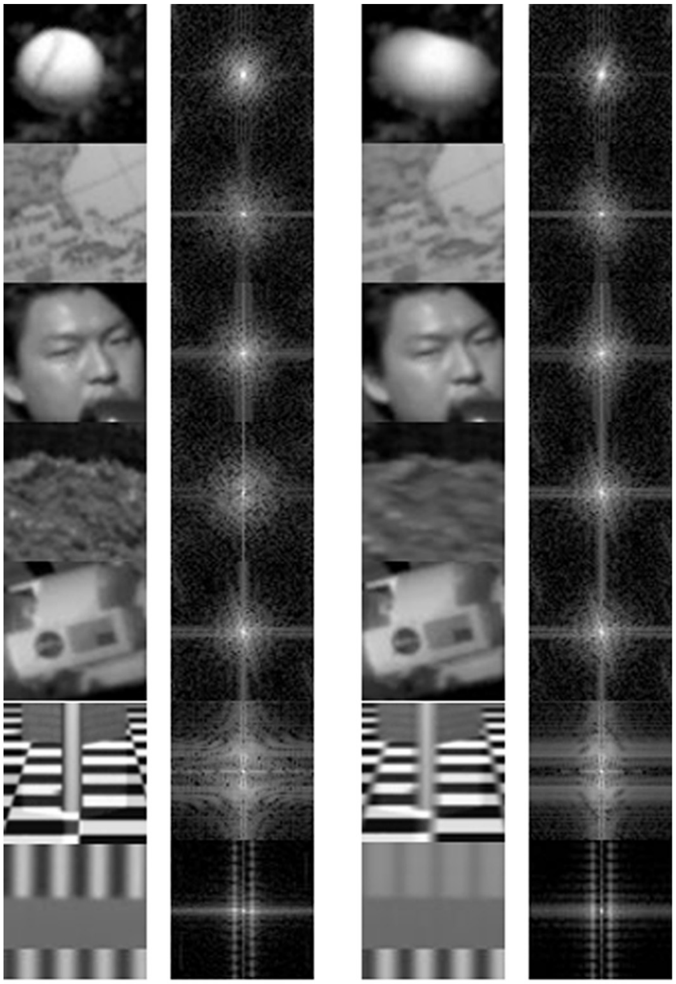
taken by a high-speed camera (nac Imaging Technology, MEMRECAM) at 1000 fps with an open (1/1000 sec) shutter. The images at 500, 333, 250, 125, and 62 fps were produced by frame averaging. The Ball, Car, and Garden were taken in daylight, while the Earth, Space, and Screen mover were taken in metal-halide light (ARRI Daylight Compact 1200) in the non-flicker mode. Computer graphics was generated and frame averaging was applied to a sequence of frames generated corresponding to 1000 fps. The white lines in each figure indicate the attention region which we defined and also the calculation point of the motion vector which was used for analysis. These lines were not shown to the subjects. The image size, display time, and number of colors were limited by the visual stimulus generator (VSG, 32 MB, 256 tables of 8 bits per color). The display unit was a 21-in. customized CRT monitor (Sony Multiscan G520) which had 480 Hz of vertical-frequency capability. The decay time from 90 to 10% of the green and blue phosphors of the CRT was 0.3 msec and of red was 1.0 msec. The images made for 500 fps were displayed at 480 Hz, which is the maximum frequency of the display unit. The images made for 500 fps are referred to as 480 fps hereafter. The 480-fps stimulus image and a randomly selected test image were shown to the subjects three times. In the experiment to evaluate blur, the subjects were instructed to move their head and eyes freely, whereas in the experiment to evaluate jerkiness they were instructed to fixate on a marker in the center of the image area. The subjects were required to answer the degradation of the test motion image compared to the 480-fps motion image. The subjects consisted of five adults. Two experimental runs were applied. The viewing distance was 45 cm in order to set the viewing angle to more than 30°. The lighting condition was under a topical fluorescent light at 200 lux at the CRT face. The subjective evaluation values of the test image were as follows:

5: imperceptible,
4: perceptible but not annoying,
3: slightly annoying,
2: annoying,
1: very annoying.

These levels correspond to quality values as follows:

5: excellent,
4: good,
3: fair,
2: poor,
1: bad.

The evaluation value 4.5 is the perception limit in the EBU recommendation. According to the results in Figs. 16–18, 120 fps provides good improvement compared to 60 fps and the evaluation values of pursuit viewing with open shuttered images and fixation viewing with open shuttered images are saturated near 250 fps and reaches the perception limit of both blur and jerkiness. And evaluation values of fixation viewing with 1/500-sec shuttered images were saturated near 250 fps except for a test motion image (Garden).



$$F(k_1,k_2)=\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} f(n_1,n_2)W_1^{n_1k_1}W_2^{n_2k_2},\ \begin{cases}k_1=0,1,\cdots,N_1-1\\ k_2=0,1,\cdots,N_2-1\end{cases} \quad (1)$$

$$\text{where}\quad W_1=e^{-i2\pi/N_1},\ W_2=e^{-i2\pi/N_2}$$

$$F_{amp}(k_1,k_2)=|F(k_1,k_2)|=\sqrt{\mathrm{Re}\{F(k_1,k_2)\}^2+\mathrm{Im}\{F(k_1,k_2)\}^2} \quad (2)$$

$$F_{ave}=\Big[\sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1}\{F_{amp}(k_1,k_2)\cdot\sqrt{k_1^2+k_2^2}\}\Big]\Big/\sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1}F_{amp}(k_1,k_2) \quad (3)$$

$$F_{ratio}=F_{ave\_62fps}/F_{ave\_480fps} \quad (4)$$

FIGURE 19 — (a) Stimulus images and FFT amplitude images. (b) Equation of the ratio of averaged Fourier frequency.

4 Dependence in the results of subjective evaluation on spatial frequency and motion speed of the visual target

Figure 19(a) shows a pair of stimulus images and their Fourier amplitude images at 480 and 62 fps. Figure 19(b) shows the equation for the ratio of averaged Fourier frequency, F_{ratio}, where $f(n_1,n_2)$ is the image of $n_1 \times n_2$ pixels, $F(k_1,k_2)$ is the discrete Fourier transform, and F_{amp} is the Fourier amplitude and we defined F_{ave} as the averaged Fourier frequency. F_{ave} corresponds to the richness of the high spatial frequency of the image and blur of 62–480 fps corresponds to F_{ratio}. Figure 20 shows the 62-fps evaluated value of blur against the logarithm of the ratio of the averaged Fourier (spatial) frequencies at 62 and at 480 fps. Figure 21 shows the 62-fps evaluated value of blur against the logarithm of the motion vector magnitude. The dependence is linear except in the case of the Ball. The maximum magnitude of the motion vector of the Ball was 78 dps which is

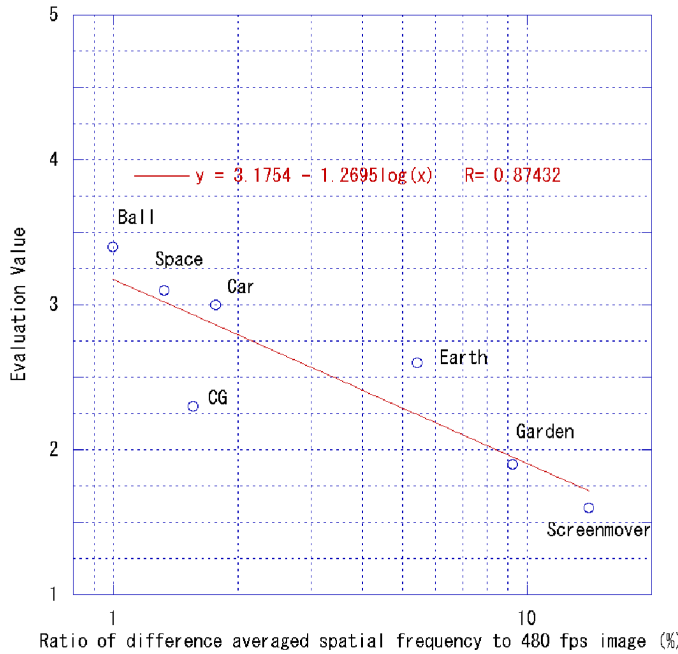


FIGURE 20 — Evaluated value of 62-fps blur *vs.* ratio of averaged Fourier (spatial) frequencies at 62 fps and at 480 fps.

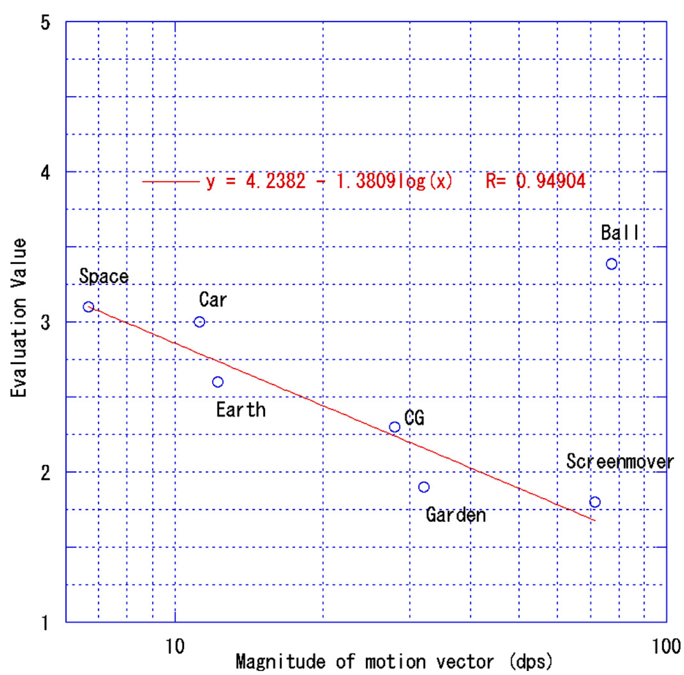


FIGURE 21 — Evaluated value of 62-fps blur against "magnitude of motion vector."

far beyond the resolution limit of 60 dps which we obtained in our first experiment, and the display time was 0.52 sec; therefore, appropriate evaluation may not have been possible. The other values of the maximum magnitude of the motion vector and the display time are as follows:

Space: 7 dps, 2.0 sec
Car: 11 dps, 1.32 sec
Earth: 12 dps, 2.0 sec
CG: 23 dps, 2.0 sec
Garden: 32 dps, 2.0 sec
Screen mover: 72 dps, 2.0 sec.

5 Discussion

Because there is a large difference between the visual resolution limits for the real and the displayed images in Fig. 6, improvement of the temporal resolution of hold-type displays is required for the realization of higher-quality motion images. The perceived MTF of hold-type displays has been extensively studied.[5] The decrease of visual resolution of hold-type displays in smooth pursuit even for images taken with a high-speed shutter is caused mainly by retinal slip, which today is well-known. For example, when we see a moving object on a hold-type display with a vertical sync of 60 Hz, we perceive blur corresponding to 1/60 sec caused by slipping of the retinal image as the eyes follow the trajectory of movement of the object smoothly, although the image maintains its position on the display during the frame time. A high-speed shutter and impulse-type display reduce such perception of blur, but images displayed at a common frame rate such as 60 fps often cause jerkiness, especially for a fixation view. A high frame rate, therefore, is important to improve the picture quality by minimizing both blur and jerkiness. Figures 16–18 show that the perception limit for blur and jerkiness is around 250 fps. When we consider the 24-fps frame rate used by cinema, 240 fps is an ideal frame rate because it provides the highest motion quality with straightforward image processing, such as synthesizing natural images and computer graphics, because 240 is a multiple of 24. Some sample images simulated for 60 and 240 fps are shown for computer graphics in Fig. 22 and for natural images in Fig. 23. In Fig. 22, each ball has a different motion vector and (a) and (b) are for the case of the pursuit of the 8 ball and (c) and (d) are for the case of the fixation view. In both cases, 240 fps provides higher quality compared to that at 60 fps. In Fig. 23, blur and jerkiness are seen in (a) and (b) at 60 fps but (c) at 240 fps provides higher quality.

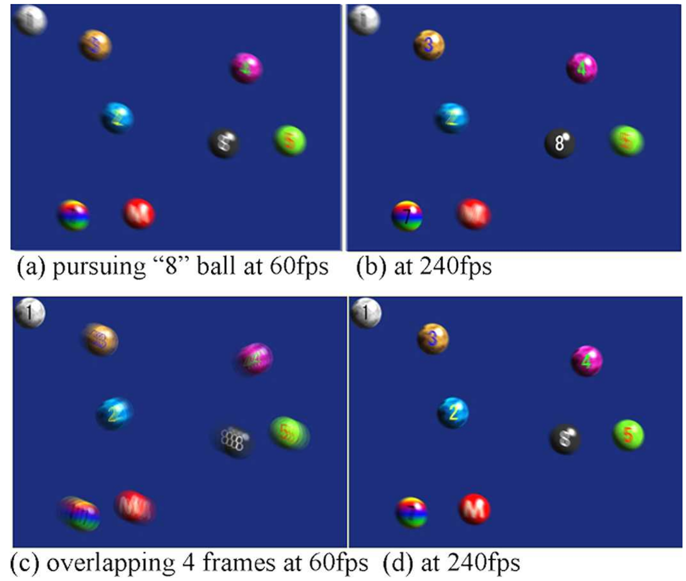


FIGURE 22 — Comparison between 60- and 240-fps computer-graphics image (simulated).

6 Conclusion

The frame rate required to eliminate both blur and jerkiness is around 250 fps. Considering the 24-fps frame rate of cinema, 240 fps is the ideal frame rate which provides the highest quality of motion images and straightforward image synthesizing of natural and computer-graphics images.

Even a 120-fps image is improved compared to 60 fps images with respect to the elimination of blur and jerkiness.

The subjective evaluation of blur has a linear dependence on the logarithm of the averaged Fourier (spatial) frequency and also on the logarithm of the magnitude of the motion vector of the attention region of the image.

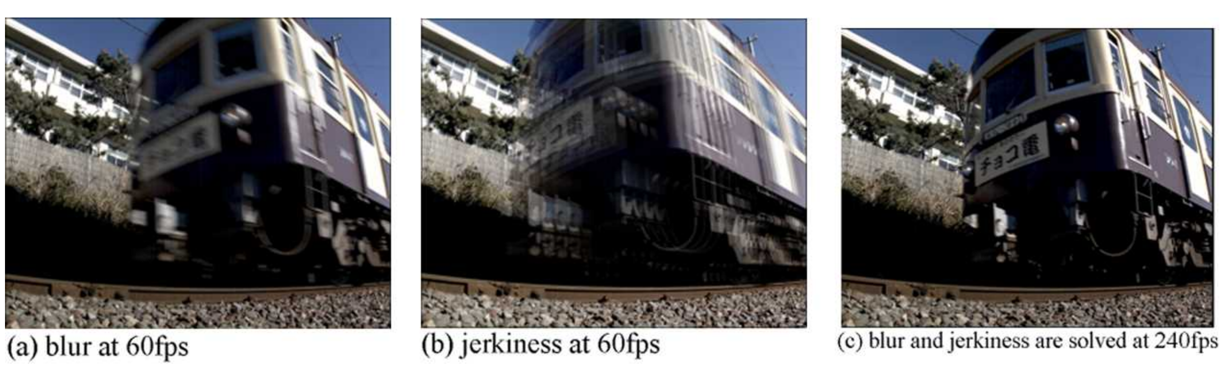


FIGURE 23 — Comparison between 60- and 240-fps natural image (simulated).

7 Acknowledgments

The authors would like to thank Dr. Ishimoto, Sony PCL, and colleagues at Sony for useful discussions and help with the subjective evaluations.

References

1 Rec. ITU-R BT.500-7, "Methodology for the subjective assessment of the quality of television pictures" (1995).
2 G Westheimer, "Eye movement responses to a horizontally moving visual stimulus," *A.M.A. Arch Ophthal* **52**, 932–941 (1954).
3 C H Meyer, A G Lasker, and D A Robinson, "The upper limit of human smooth pursuit velocity," *Vision Res* **25**, No. 4, 561–563 (1985).
4 ANSI T1.801.02-1996.
5 T Kurita, A Saito, and I Yuyama, "Consideration of perceived MTF of hold-type display for moving images," *IDW '98* **3D3-4**, 823–826 (1998).

Yoshihiko Kuroki received his B.Eng. degree from Keio University and joined Sony in 1979. He was a visiting scientist at MIT from 1993 to 94 in the Department of Mechanical Engineering. Previously, he had been engaged in developing broadcasting cameras, and he was a head of the Vision and Cognition Laboratory of the Research Center. He is currently a senior researcher in the TV & Video Business Group and a Senior Manager of the Technology Development Group of Sony Corp. He is a member of SID.

Tomohiro Nishi received his M.Eng. degree from Tokyo Institute of Technology in 1998. He joined Sony Corp. in 1998. At present, he is engaged in the research and development of image-processing algorithms for displays.

Seiji Kobayashi received his M.Eng. degree from Seikei University. He began his professional career in 1991 by joining Sony Corp. He has been developing broadcast video equipment, 3-D computer graphics, LSIs, and applications of 3D-CG. In 1999, he joined Sony-Kihara Research Center, Inc., and developed image-processing algorithms. Since April 2006, he has been a senior manager at the Technology Development Group of Sony Corp.

Hideki Oyaizu received his M.Eng. degree from Tohoku University. Since he joined Sony Corp. in 1989, he has been developing visual inspection systems and image-processing algorithms for industrial use. In 1999, he joined Sony-Kihara Research Center, Inc., and developed 3-D data-capturing camera systems and video-processing algorithms.

Shinichi Yoshimura received his M.Eng. degree from Kyoto University. Since he joined Sony Corp. in 1987, he has been developing visual-inspection systems, image-processing algorithms, and 3-D image-sensing systems for industrial use. From 1991 to 1993, he was a visiting research scientist at the Robotics Institute of Carnegie Mellon University. In 1999, he joined Sony-Kihara Research Center, Inc., and developed intelligent CMOS image sensor and its application systems. Since April 2006, he is a Senior Manager at the Technology Development Group of Sony Corp. He is a member of IEEE.