Control of subjective depth on 3-D displays by a quantified monocular depth cue

Shuichi Takahashi (SID Member)
Takanori Ishikawa
Yasuhide Hyodo
Isao Ohashi
Yoshihide Shimpuku (SID Member)
Kazuya Matsubara
Kazumichi Matsumiya
Satoshi Shioiri

Abstract — This paper considers the architecture and the effectiveness of new algorithms that control the subjective depth on 3-D displays by modulating the contrast of stereoscopic pictures on the basis of a quantified monocular depth cue. First, a psychophysical experiment to quantify the relationship between contrast and subjective depth was conducted. This experimental result shows that the higher the contrast, the nearer the object will be perceived, corresponding to the qualitative relationship. Second, this result was applied to image-processing algorithms that expand or contract the subjective depth of stereoscopic pictures. Subjective assessments to verify the effectiveness of the algorithms were also conducted. The results suggest that the algorithms will allow viewers to experience a highly realistic sensation.

Keywords — *3-D display, subjective depth, monocular-depth cue, contrast, stereoscopic picture.*

DOI # 10.1889/JSID19.1.1

1 Introduction

Observing the real world, human beings will perceive depth on the basis of various depth cues. The depth cues can be classified into binocular depth cues, such as binocular disparity and convergence, and monocular depth cues, such as aerial perspective and occlusion. Being one of the most effective cues in perceiving depth, binocular disparity can arouse a feeling of highly realistic and dramatic sensation. As a result, some pictures have greater binocular disparity to attract an audience. However, some studies have also revealed that excessive binocular disparity causes visual fatigue or sickness.[1] To avoid these symptoms, the binocular disparity should be reduced so that most viewers feel comfortable while watching stereoscopic pictures, but those with limited binocular disparity sometimes fail to attract observers.

We have considered monocular depth cues to be effective means in order to overcome this contradiction. The cues include aerial perspective and occlusion. In particular, aerial perspective has often been used as a pictorial technique in the *Mona Lisa* and other paintings. This comprises several factors such as contrast, spatial frequency, and saturation. A previous study on the relationship between contrast and subjective depth[2] shows that the higher the contrast between an object and its background in a picture, the nearer the object is perceived to a viewer. This relationship suggests that the technique of aerial perspective can be applied to control of subjective depth.

The purpose of this paper is to examine the architecture and the effectiveness of our algorithms which control the subjective depth on 3-D displays by modulating the contrast of stereoscopic pictures on the basis of a quantified monocular depth cue. First, results of psychophysical experiments for the quantification are described. Second, algorithms with the quantified relation are explained. Third, equipment and methods used for a subjective assessment are mentioned. Pictures processed by the proposed algorithms are also presented. Finally, results of the assessment are shown and discussed.

2 Quantification of depth cue

To quantify the relationship between the contrast of a picture and the subjective depth, we had conducted some psychophysical experiments.[3,4] The results supported the finding that the higher the contrast is between an object and its surrounding background in a picture, the nearer the region is perceived to a viewer.[2] Figure 1 shows the quantified relationship between an effective contrast and a point of subjective equality (PSE). The effective contrast is the ratio of the actual contrast of a picture to the threshold of the contrast sensitivity. This sensitivity was obtained by our psychophysical experiments at spatial frequencies of 1, 2, 4, and 8 cycles per degree. The PSE here is (CPD) translated into the binocular disparity, and the unit of the PSE is minutes of arc. This PSE can be converted into subjective depth derived from the contrast of a picture by calculating the geometry with a pixel pitch of a display, a viewing distance, and an inter-pupillary distance.

Two logarithmic functions can be used for the approximation of the quantified relationship. One is used in the spatial frequencies equal to or lower than 2 cpd, and the other is used in the spatial frequencies higher than 2 cpd. The logarithmic functions are defined here by

$$D_{\mathrm{low}} = 49.3 * \log_{10}(C_{\mathrm{eff}}) - 26.7, \tag{1}$$

$$D_{\mathrm{high}} = 30.0 * \log_{10}(C_{\mathrm{eff}}) + 4.2, \tag{2}$$

where D_{low} and D_{high} are the approximate subjective depths in the spatial frequencies equal to or lower than 2 cpd and

Based on a paper presented at the 2010 SID Symposium held May 25–28, 2010, in Seattle, Washington, USA.
S. Takahashi, T. Ishikawa, Y. Hyodo, I. Ohashi, and Y. Shimpuku are with Sony Corp., Department of VSD, RFVS, CDDG, Goetenya TEC 5G3F, 5-1-12 Kita-Shinagawa, Shinagawa, Tokyo 141-0001, Japan; telephone +81-3-5448-4525, e-mail: Shuichia.Takahashi@jp.sony.com.
K. Matsubara, K. Matsumiya, and S. Shioiri are with Tohoku University, Sendai, Miyagi, Japan.

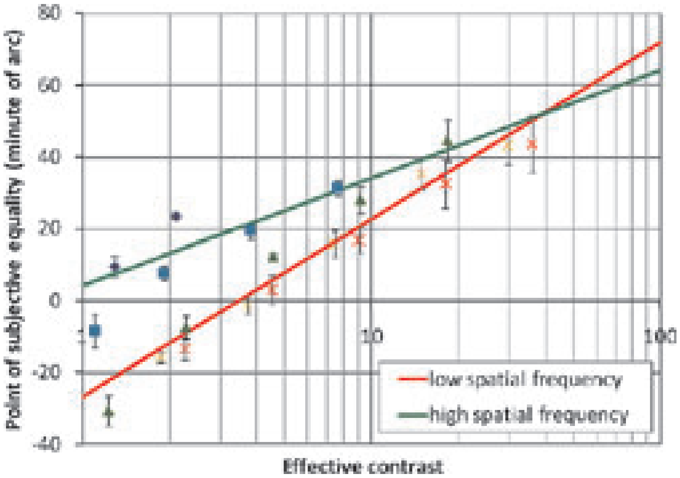


FIGURE 1 — Subjective depth defined by an effective contrast of visual stimuli. The horizontal axis shows the effective contrast. The effective contrast is defined as the ratio of the actual contrast of a picture to the threshold of the contrast sensitivity. The vertical axis shows the point of subjective equality, which means the subjective depth derived from the contrast of an image. Positive values mean the stimuli are perceived nearer. The red and green lines indicate the functions for spatial frequencies equal to or lower than 2 cpd and higher than 2cpd, respectively.

higher than 2 cpd, respectively, and C_{eff} is the effective contrast. For example, Eq. (1) shows that observers will perceive a 1-cpd visual stimulus with the effective contrast of 10 at the position where an object with a binocular disparity of 22.6 arcmin is located.

3 Image-processing algorithm

3.1 Overview

Image-processing algorithms based on the shown quantified relationship were created. The algorithms consist of two main ideas: depth expansion and depth contraction. The depth-expansion algorithm is used to expand the subjective depth derived from the contrast of a picture, and the depth-contraction algorithm is used to contract the subjective depth.

Figure 2 is a block diagram of the image-processing algorithms which were applied to pictures. Input and output pictures were stereoscopic. Both left- and right-eye pictures were processed through six steps as follows: luminance calculation, contrast analysis, frequency analysis, depth estimation, modulation-amplitude calculation, and contrast modulation.

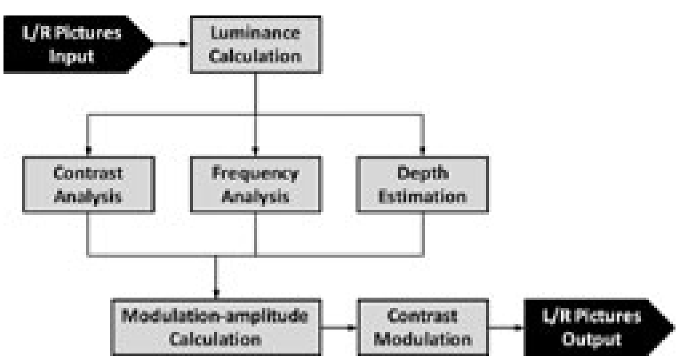


FIGURE 2 — A block diagram of image processing.

3.2 Luminance calculation

To calculate the contrast of the pictures, the luminance of each input picture was calculated on the basis of a relation defined in ITU-R BT.709.[5] All the pictures presented to observers were converted into luminance pictures in this step because the effect of hue and saturation of the pictures should be eliminated.

3.3 Contrast analysis

The contrast of the luminance picture was analyzed using the calculated luminance. The contrast is defined here as the Michelson contrast by

$$C_M = (Y_{max} - Y_{min})/(Y_{max} + Y_{min}), \tag{3}$$

where C_M is the Michelson contrast, and Y_{max} and Y_{min} are the maximum and the minimum values of luminance in one region of the picture, respectively. Ideally, the contrast is analyzed by every pixel, but here we analyzed it by 2 × 2 pixels to accelerate the calculation speed. This did not cause any perceptible artifacts. Since the Michelson contrast is proportional to the effective contrast according to the definition of the effective contrast, the functions described in Eqs. (1) and (2) are also valid for the contrast modulation based on the Michelson contrast.

3.4 Frequency analysis

Spatial frequencies of the luminance picture were analyzed with the response to Gabor filters of 1, 2, 4, and 8 cpd. The sizes of the filters were set to 56 × 56, 28 × 28, 14 × 14, and 7 × 7 pixels, respectively, in consideration of the viewing distance of the picture in a subjective assessment described later. Results of the convolution between pixel values of the picture and the Gabor filters describe the intensity of each spatial frequency and reveal how much the picture contains the four specific spatial frequencies.

3.5 Depth estimation

The horizontal distance between a region in a left-eye picture and the corresponding region in a right-eye picture was analyzed in order to estimate the depth of the region in the pictures. In this analysis, every corresponding region in the pictures can be detected by a block-matching-based algorithm, and then the horizontal distance can be estimated. The horizontal distance can be converted to the depth of the regions by calculating the geometry with a pixel pitch of a display, a viewing distance, and an inter-pupillary distance. When the horizontal distance is larger than zero, for exam-

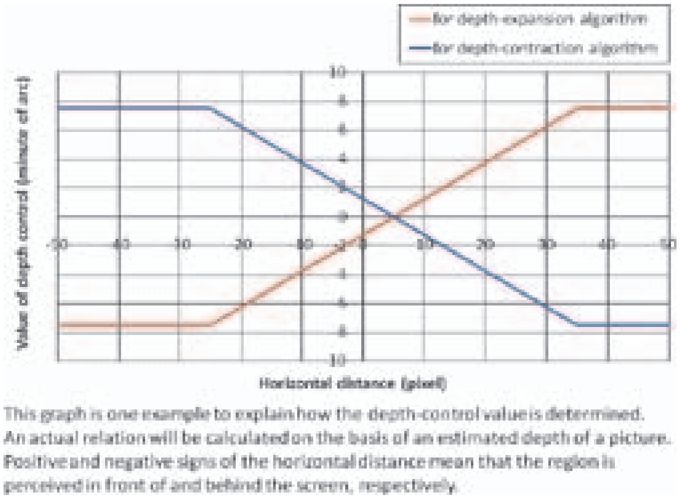


FIGURE 3 — The depth-control value. This relation is one example to explain how the depth-control value is determined. An actual relation will be calculated on the basis of an estimated depth of a picture. Positive and negative signs of the horizontal distance mean that the region is perceived in front of and behind the screen, respectively.

ple, the region will be perceived further from the display screen. The information obtained here is used in the following step.

3.6 Modulation-amplitude calculation

The amplitude of contrast modulation was calculated using a depth-control value determined in advance and the quantified relationship shown in Fig. 1. The depth-control value determines how much the subjective depth of regions in a picture will be shifted. Here, it was set within ±7.5 minutes of arc. Figure 3 indicates one example to explain how the depth-control value is determined. The actual value will be calculated on the basis of estimated depth of a picture.

As shown in Fig. 3, the tendency of the value depends on the algorithm selected for image processing. The value was proportionally changed according to the horizontal distance analyzed in the previous step. This means that the nearest region in a depth-expanded picture will be perceived nearer to observers than that of an original by 7.5 minutes of arc; the farthest region will be perceived further from observers by 7.5 minutes of arc; and a region whose horizontal distance is +15 pixels will be perceived further from observers by 2.5 minutes of arc. By proportionally changing the depth-control value, the relative depth position of all regions in a picture can be maintained. This value can be arbitrarily determined in consideration of processing effectiveness and overall picture quality.

After the depth-control value was determined, the actual amplitude of contrast modulation was calculated using the value and the quantified relationship described in Eqs. (1) and (2). To determine which equation should be selected, spatial frequencies contained at a processed region in a picture were examined. This information was already obtained by the analysis described in Sec. 3.4. Although all the frequency elements should have been considered because pictures usually contain various spatial frequencies, we selected one dominant frequency among the four (1, 2, 4, and 8 cpd) for simplicity by comparing the intensity of each spatial frequency. In other words, even if the processed region has the highest intensity of 4 cpd and also contains various spatial frequencies, the contrast of the region will be modulated across all the contained spatial frequencies on the basis of Eq. (2).

The following is one calculation example of how to change the subjective depth of a region in a picture with our depth-expansion algorithm. Given that one region with the spatial frequency of 1 cpd, the Michelson contrast of 10 and a horizontal distance of 15 pixels is located in an original picture, how much will the contrast of the region be modulated? Because its spatial frequency is lower than 2 cpd, Eq. (1) is selected. And Fig. 3 shows that the depth-control value is set to 2.5 minutes of arc at a horizontal distance of 15 pixels. The amplitude of contrast modulation in this region will be calculated by solving the following equations:

$$D_{\mathrm{low\,(org)}} = 49.3 * \log_{10}(C_{\mathrm{eff\,(org)}}) - 26.7, \tag{4}$$

$$D_{\mathrm{low\,(mod)}} = 49.3 * \log_{10}(C_{\mathrm{eff\,(mod)}}) - 26.7, \tag{5}$$

$$D_{\mathrm{low\,(mod)}} - D_{\mathrm{low\,(org)}} = 2.5 \text{ (minutes of arc)}, \tag{6}$$

where $D_{\mathrm{low\,(org)}}$ and $D_{\mathrm{low\,(mod)}}$ are the subjective depth of the region in the original picture and that of the contrast-modulated one, respectively, and $C_{\mathrm{eff\,(org)}}$ and $C_{\mathrm{eff\,(mod)}}$ are the effective contrast in each picture. Therefore, as shown in Eq. (7), the effective contrast of the region in the contrast-modulated picture should be 1.12 times higher than that of the original one.

$$C_{\mathrm{eff\,(mod)}}/C_{\mathrm{eff\,(org)}} = 1.124. \tag{7}$$

Since this effective contrast is defined here as the ratio of the actual contrast of a picture to the threshold of the contrast sensitivity, the Michelson contrast should also be heightened 1.12 times. The amplitude of contrast modulation in all the regions of the original picture was repeatedly calculated, and then the contrast of the picture was modulated. This amplitude was calculated by 2 × 2 pixels to accelerate the computation speed and this did not cause any perceptible artifacts.

With regard to our depth-contraction algorithm, a similar procedure can be adopted; however, the depth-control value has a different tendency because the subjective depth of a picture will be contracted.

3.7 Contrast modulation

The contrast of the original picture was modulated on the basis of the modulation amplitude calculated above. First, an original luminance picture was analyzed by the discrete Fourier transform (DFT) to calculate intensity in the spatial-frequency domain. Phase information obtained here was not used for this contrast modulation. The window size of the DFT was set to 56 × 56 pixels and an analysis unit was

2×2 pixels to accelerate the computation speed. These four pixels were located at the center of the window. The entire picture was analyzed with this window shifted by every two pixels and lines.

Second, in order to modulate the contrast of the original picture, the intensity calculated above was multiplied by the modulation amplitude at the center four pixels. These center pixels had the same Fourier coefficient because they had the same intensity and the same amplitude. As mentioned in Sec. 3.6, the intensity of all spatial frequencies was increased or decreased on the basis of either Eq. (1) or Eq. (2). Which equation will be selected depends on the dominant spatial frequency at the center pixels. In this step, the contrast needs to be heightened or lowered in order to keep the average luminance of the picture constant. Therefore, after the Fourier coefficients were calculated, only the AC components of spatial frequencies were modulated so that the Michelson contrast became the target value, but the DC component was not changed.

Finally, all the Fourier coefficients in the frequency domain were converted into pixel values by the inverted DFT.

4 Subjective assessment

4.1 3-D display systems

We conducted a subjective assessment using two stereoscopic 3-D displays and a pair of liquid-crystal shutter glasses. The displays were 40-in. LCDs with a resolution of 1920×1080 and a frame rate of 240 Hz.

Figure 4 describes the display systems used for this assessment. All the original and processed pictures were stored in this picture source. At this control board the pictures were converted to a video-signal format appropriate for the displays. A synchronous signal which can operate the liquid-crystal shutter glasses was also generated in this control board. This synchronous signal was transmitted to this synchronous signal emitter placed near the displays. Video signals were directly input into panels of the displays to bypass all the image processing steps unrelated to our algorithms.

As shown in Fig. 4, left- and right-eye pictures were displayed on the screen twice in a row to reduce the influence of cross-talk, and both pictures were alternately displayed in synchronization with the switching of the liquid-crystal shutter glasses. The glasses, operating at 60 Hz, enabled the displays to deliver 3-D functionality without any loss of resolution and a flicker. The duty ratio of the glasses was set to 12.5% in order to suppress cross-talk under 3%.

4.2 Apparatus

The experimental booth was dark. The displays were placed on a table next to each other. The distance between an observer and the displays was 1.5 m, which was approximately three times the height of the display screens. The visual angles to

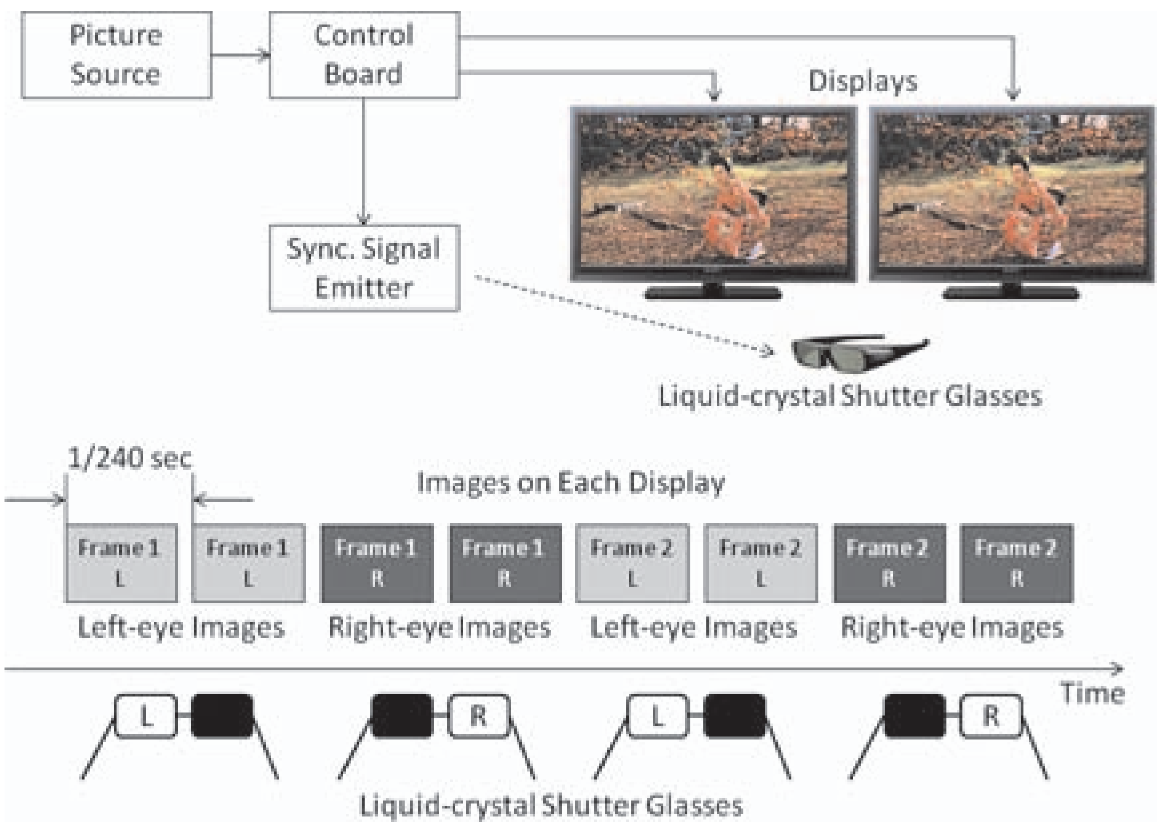


FIGURE 4 — Display systems: connections between the displays and other equipment, and the operations of the displays and the shutter glasses.

FIGURE 5 — Test materials used for this assessment.

the screens were 33° and 19° in the horizontal and vertical meridians, respectively.

4.3 Observers

A total of 25 observers, 23 males and 2 females, participated in this assessment. They were aged from the 20s and the 50s. All had normal stereo vision confirmed by viewing stereoscopic pictures. Prior to the assessment, we adequately explained objectives and a procedure to them.

4.4 Test materials

The six pictures shown in Fig. 5 were used as test materials. These pictures include deep and/or shallow depths, and near and/or far regions. Pictures A, B, and C are stereoscopic 3-D still and motion pictures standardized by ITU-R BT.1438.[6] Pictures D, E, and F are taken from frames of stereoscopic 3-D motion pictures by NHK Media Technology. One frame of every motion picture was selected and converted from an interlaced format into a progressive bitmap file by an IP conversion technique.

TABLE 1 — Five-scale evaluation terms used for this assessment.

+2	Very deep
+1	Deep
0	Same
-1	Shallow
-2	Very shallow

Two types of processed pictures were generated with our algorithms. One was a depth-expanded picture and the other was a depth-contracted picture. A total of 18 pictures, including original ones, were prepared for this assessment.

4.5 Procedure

We assessed the subjective depth of the pictures perceived by observers using a paired-comparison method. First, two pictures among one original and two processed pictures were randomly selected and displayed on screens. Second, observers viewed the pictures with no time limitations. And then they judged the depth of the picture on the right screen with reference to that of the left. Finally, they told a score based on five-scale evaluation terms listed on Table 1. Gray pictures were displayed for 5 sec on both screens when paired pictures were switched.

5 Results and discussion

5.1 Processed pictures

Figure 6 shows all the original and processed pictures used in this experiment and the normalized amplitude of contrast modulation. Pictures of modulation amplitude in this figure represent the normalized amplitude of contrast modulation in test materials. Here, brighter pixels in these pictures indicate the higher modulation amplitude.

Nearer and further regions in the depth-expanded pictures become clearer and a little blurred, respectively, because the depth-expansion algorithm heightens the contrast of nearer regions and lowers that of further ones. On the other hand, nearer and further regions in the depth-contracted pictures become a little blurred and clearer because the depth-contraction algorithm lowers the contrast of nearer regions and heightens that of further ones.

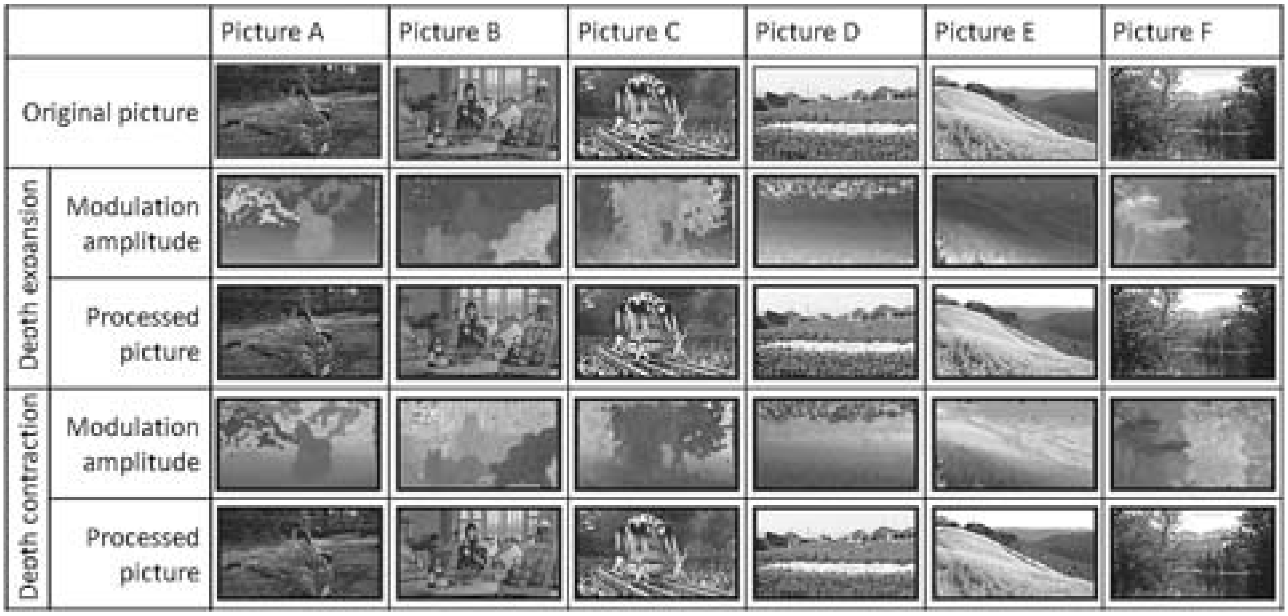


FIGURE 6 — Original and processed pictures, and normalized amplitude of contrast modulation shown as pictures. Original luminance pictures were converted from the pictures shown in Fig. 4. Original pictures were processed by the depth-expansion or the depth-contraction algorithm. In the modulation-amplitude pictures, brighter pixels have higher modulation amplitude.

5.2 Psychological measure

In order to evaluate the effectiveness of the algorithms, a psychological measure of subjective depth was calculated with the results of the assessment. Figure 7 shows the psychological measure of subjective depth for all the observers. The horizontal axis shows the test materials and the vertical axis shows the measure in each picture. As shown in Fig. 7, it is fair to say that out depth-expansion and depth-contraction algorithms can affect the observers' subjective depth. This is because this psychological measure of each depth-expanded picture was larger than that of each original picture and because the measure of each depth-contracted picture was smaller than that of each original picture. We also analyzed whether or not the processed pictures significantly differed from their original ones to verify the effectiveness of the algorithms. The result shown in Fig. 7 indicates that all the processed pictures except for the depth-expanded picture B had significant difference from the original ones under the significance level of 5%.

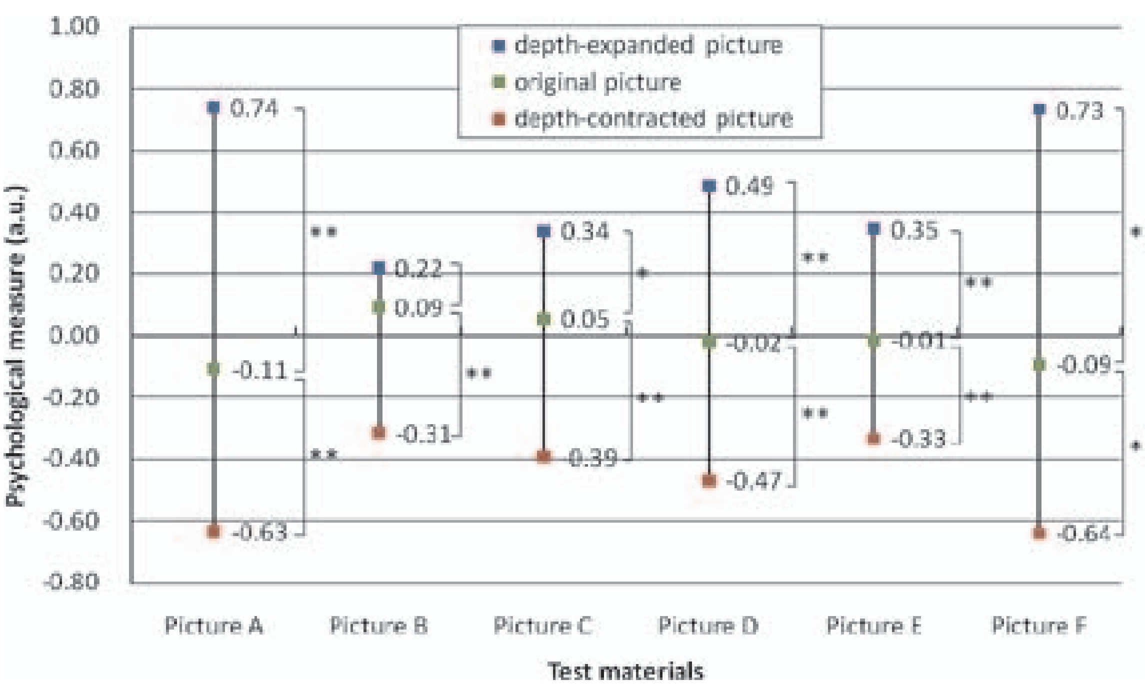


FIGURE 7 — The psychological measure for all the test materials. The horizontal axis shows the test materials and the vertical axis shows the psychological measure in each picture. Asterisks indicate the significant difference between pictures.

5.3 Discussion

These results suggest that the proposed algorithms can clearly control the subjective depth of pictures in the depth-expansion and depth-contraction directions. Since this effect can be produced without changing binocular disparity, the algorithms allow people to feel a highly realistic and dramatic sensation if they are applied to pictures with limited binocular disparity.

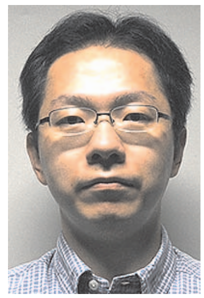
The effectiveness of the algorithms was confirmed; however, its dependence on pictures should be reduced. We hypothesized on two factors which will lead to the result that the depth-expanded picture B had no significant difference. The first factor is disparity distribution of the picture. The result of the depth estimation reveals that teddy bears and a chair, which had large binocular disparity, would be perceived in front of the screen. This might cause some observers to find the picture unnatural and difficult to see. Additionally, since these regions were located close to the edge of the display, the subjective depth was forced to be distorted. This phenomenon might cause observers to perceive unstable depth. This hypothesis was confirmed by the fact that many of the observers had some difficulty in viewing the nearest regions. The second factor is an actual depth which the picture B contains. The actual depth between the nearest (the teddy bears and the chair) and the farthest regions (windows and walls) in the picture B is considered 5–10 m, while the actual depths of other pictures seem to be tens to thousands of meters. At present, the actual depth between regions cannot be detected because the current depth-estimation technique estimates only the relative depth of regions in a picture. This means that the contrast of the nearest region is heightened and that of the furthest is lowered, even though the actual depth between two regions is only 1 m. This picture will seem completely different from what we see in the real world. Some observers pointed out that the depth-expanded picture B seemed unnatural or strange because the windows and the walls, which were located near the lady sitting on a sofa, were perceived more blurred than what they are. This suggests that the actual depth contained in pictures should be taken into consideration in the contrast modulation.

6 Conclusion

We applied one quantified monocular depth cue to image-processing techniques, and then assessed the subjective depth of stereoscopic 3-D pictures processed by the depth-expansion and the depth-contraction algorithms. The statistically analyzed results revealed that the algorithms can effectively control the subjective depth. Therefore, it is suggested that these algorithms will allow observers to experience a highly realistic and dramatic sensation even when they view pictures with limited binocular disparity. We also confirmed the dependence of the effectiveness on pictures and hypothesized on the dependence factors. These hypotheses probably indicate that disparity distribution of a picture and actual depth between regions should be considered to make our algorithms more effective.

References

1 S. Yano *et al.*, "A study of visual fatigue and visual comfort for 3D HDTV/HDTV images," *Displays* **23**, 191–201 (2002).
2 M. Farnè, "Brightness as an indicator to distance: Relative brightness per se or contrast with the background?," *Perception* **6**, 287–293 (1977).
3 K. Matsubara *et al.*, "The effect of luminance contrast and stimulus eccentricity on the subjective depth," *IEICE Tech. Report* **109**(469), 71–75 (2010).
4 K. Matsubara *et al.*, "The effect of luminance contrast and stimulus distance on the subjective depth," *Vision* **22** (supplement for APCV2010), 76, (2010).
5 "Rec. ITU-R BT.709-5: Parameter values for the HDTV* standards for production and international programme exchange" (2002).
6 "Rec. ITU-R BT.1438: "Subjective assessment of stereoscopic television pictures" (2000).



Shuichi Takahashi received his B.E. and M.E. degrees in electronics, information and communication engineering from Waseda University in 1999 and 2001, respectively. He now works for Sony Corp. He has years of experience in 3-D technologies and human vision. He is a member of the Society for Information Displays (SID) and the Institute of Electronics, Information and Communication Engineers (IEICE). His research interests include display technologies, image processing, the human visual system, and cognition and perception.

Takanori Ishikawa received his B.S. and M.S. degrees in electronic engineering from Doshisha University in 1999 and 2001, respectively. Since 2001, he has been working for Sony Corp. His interests include 3-D displays and image processing with human vision.

Yasuhide Hyodo received his B.S. degree in information engineering from the Tokyo Institute of Technology in 2007 and received his M.S. degree in information and communication engineering from the University of Tokyo in 2009. In 2009, he joined Sony Corp. and he is currently working in the area of 3-D systems development. His research interests include 3-D displays, visual perception, image processing, and computer vision.

Isao Ohashi received his B.S. degree in electronic engineering from Ibaraki University in 1982 and joined Sony Corp. He is now Deputy General Manager of the Visual Systems Development Department, Sony Corp.

Yoshihide Shimpuku is a candidate for a Ph.D. degree in physics. He is General Manager of the Visual Systems Development Department, Core Device Development Group, at Sony Corp.

Kazuya Matsubara received his Ph.D. degree in engineering from Chiba University in 2006. Then, he was a post-doctoral researcher at Tokyo University until March, 2008. From April, 2008 to March, 2009, he was a post-doctoral researcher at Kanazawa University. Since 2009, he has been a post-doctoral researcher at the Research Institute of Electrical Communication, Tohoku University. His research interests include visual psychophysics, depth perception, visual attention, and eye movement.

Kazumichi Matsumiya received his Ph.D. degree from the Tokyo Institute of Technology in 2000. Then, he was a post-doctoral researcher at York University in Canada until December, 2001. From January, 2002 to December 2003, he was a post-doctoral researcher at the Imaging Science and Engineering Laboratory, Tokyo Institute of Technology. From January, 2004 to March, 2005, he was a research fellow at the Human Information Sciences Laboratory, ATR. He moved to the Research Institute of Electrical Communication, Tohoku University, as a research associate in April, 2005. He is currently working there as an assistant professor. His research interests are in visual psychophysics, multimodal integration, motion perception, and eye movements.

Professor Shioiri graduated from the Tokyo Institute of Technology and received his Dr. Eng in 1986. Then, he was a post-doctoral researcher at the University of Montreal until May of 1989. From June of 1989 to April of 1990, he was a research fellow at the Auditory and Visual Perception Laboratories of the Advanced Telecommunications Research Institute. He moved to Chiba University in May of 1990, where he spent 15 years as an assistant professor, an associate professor, and a professor in the Department of Image Sciences, Department of Image, Information Sciences, and Department of Medical Systems. In 2005, he moved to Tohoku University. Since then, he has been a professor at the Research Institute of Electrical Communication at Tohoku University. His research interests include motion perception, depth perception, color vision, mechanisms of visual attention and eye movements, and modeling of visual functions.