Relief-like Depth Map Generation Algorithm

Hajime Noto, Akihiko Hashimoto, Kazuo Kimura†, and Kenji Nakazawa

Abstract

We propose a three-dimensional (3-D)-input method using a relief-like depth map generation algorithm (REAL). REAL is based on the concept of reproducing a depth map, like the relief of a coin, from 3-D input. It enables a portable 3-D-input system to be achieved using only a monocular camera and a photoflash. REAL is very simple and convenient compared with conventional technology using triangulation. In this paper, we give an overview of the concept and the depth map calculations. Results of a subjective test conducted using 3-D images generated by REAL indicate that observers perceived natural 3-D images.

1. Introduction

With the recent establishment of groups such as the 3-D Consortium [1] and the Consortium of 3-D Image Business Promotion [2], the opportunities for using three-dimensional (3-D) images are growing rapidly. 3-D-output techniques include 3-D displays based on parallax [3], [4] and lenticular [5] methods, which have been developed commercially not only for consumer products such as cellphones, but also for personal computers (PCs) and amusement machines [6]. In NTT Cyber Space Laboratories, we are researching and developing a DFD (depth-fused 3-D [7]) display that is less fatiguing to watch than conventional stereoscopic displays. Before 3-D images can find their way into consumer products, it will be necessary to resolve not only software issues such as the provision of attractive content, but also hardware issues involved in making it just as easy to input and output natural 3-D images as to take photographs with a camera.

Conventional 3-D-input technologies have mainly been for industrial applications [8], where the main requirements are accuracy and range for shape measurements. However, for consumer 3-D-input applications it is more important to concentrate on making the equipment more compact, reducing the measurement times, and eliminating the need for a specialized imaging environment—accuracy and range are not as important. Although various studies [9]-[13] have tackled these issues, they have yet to result in practical applications.

The light sectioning methods involve projecting stripes of light onto the subject so as to scan it in a number of light sections which are then imaged from another direction and converted into 3-D form by triangulation based on the direction from which the stripes were projected and the position from which the subject is viewed. Triangulation methods involving the use of projected patterns (not just the light sectioning methods) are characterized by their high accuracy. On the other hand, to detect the projected patterns correctly, it is generally necessary to use a specialized imaging environment. Moreover, since images are generally obtained one at a time while changing the projection angle of the stripe light source in order to scan the subject, the measurement time is also long. Recently there have been reports [10], [11] of an ultrafast device that can perform measurements in real time, but this device requires a high-intensity laser as the light source for the stripe lighting for adequate sensitivity.

† NTT Cyber Space Laboratories
Musashino-shi, 180-8585 Japan
E-mail: kimura.k@lab.ntt.co.jp

Stereo methods can be used to perform measurements based on the principle of triangulation from images taken by multiple cameras at different imaging positions. A characteristic of stereo methods is that measurements (subject acquisition) can be performed in real time without the need for a special imaging environment. However, since it is difficult to perform image processing to match up the same objects from within natural images taken with multiple cameras, this method generally has poor accuracy and reliability. Multi-baseline stereo methods [14] have been proposed as effective solutions for these problems, and they have been employed in several commercial stereo-method systems [15]-[17]. However, when configuring a multi-baseline stereo system it is necessary to increase the number of cameras in proportion to the number of baselines, which is undesirable from the viewpoint of compactness.

A problem shared by all methods that use triangulation is that there are some regions whose depth cannot be obtained due to occlusion. Furthermore, since the baseline must cover at least a certain length, they are not suitable for portable systems. The relationship between baseline length and imaging distance also depends on the sensor resolution, but when measurements are performed with an imaging distance of 1–2 m, the baseline should normally be at least 5–10 cm long. Since these issues affect size measurement accuracy, they are difficult to solve in principle, so it is difficult to apply measurement techniques based on the principle of triangulation to consumer applications such as 3-D input for objects at size ranges comparable to human beings.

Conversion techniques that produce 3-D images from 2-D images have also been studied [18], [19], but since these techniques do not use any information derived from measurements, they are generally unsuitable for obtaining solid images that are sufficiently natural.

This paper proposes a technique that adds pseudo-depth information, which enables it to convey a perceptually natural 3-D impression when data obtained by a rough 3-D-input method without the use of triangulation is depicted on a 3-D display. It enables a portable 3-D-input system to be achieved using only a monocular camera and a photoflash. We give an overview of the concept and the depth map calculations. Results of a subjective test conducted using 3-D images added pseudo-depth indicate that observers perceived natural 3-D images in a DFD display.

2. Concept

In this study we concentrated on objects such as coins and medals with engraved relief patterns. As shown in **Fig. 1** [20], coins often have some kind of design engraved in a relief pattern. These designs cannot be made very thick, so they are engraved with rounded forms having more or less the same height, but since human perception is based on previous learning experience, humans can perceive a 3-D impression that appears quite natural and varies with the inclination and context of the design. If a coin depicts the profile of a human face rendered with relief gradations, with a building shown in the background, then the face can be perceived in 3-D without any sense of incongruity. Thus, humans are able to perceive 3-D impressions even from thin relief images engraved on walls or coins, without experiencing any particular feelings of incongruity.

We propose an algorithm for generating relief-like depth maps as a new scheme based on the concept of generating thin relief-type 3-D bodies by building up a deformed pseudo-depth without considering the exact scale of the part to be modeled. This scheme is called REAL (relief-like depth map generation algorithm).

In REAL, an acquired image is made three-dimensional by adding relief-type distance data to it. To do this it is necessary to decide which regions of the image the relief depth information should be added to. In this paper, as a typical example of a consumer application, we assume that the algorithm is used to process pictures taken using a cellphone equipped with a camera. Earlier studies have shown that the subject is highly likely to be a human [21], [22] and that a self-portrait of a female is the most common subject [23]. This suggests that the main subject is often situated quite close to the camera. Accordingly, in the basic concept of REAL it is assumed that the parts that are to be made three-dimensional are in the

Fig. 1. Examples of relief on coins.

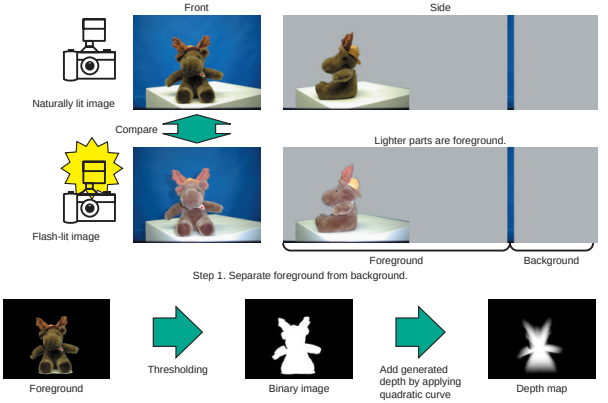


Fig. 2. Schematic diagram of REAL.

foreground. Based on this assumption, the image is roughly separated into foreground and background parts, and relief depth is added only to the foreground. Parts that are judged to be in the background are all treated as background objects regardless of their true distance. **Figure 2** summarizes this process.

2.1 Processing of REAL

A step-by-step description of REAL is presented below.

• Step 1: Separate foreground and background

First, the foreground and background are separated in order to extract the region to which the relief processing is to be applied. The difference in image intensities between a naturally lit image and a flash-lit image of the subject is evaluated and compared with a fixed threshold to obtain a binary image. The threshold value used during this differencing process is affected by numerous factors including the ambient light, the flash intensity, the distance to the subject, the reflectivity of the subject, and the signal-to-noise ratio of the camera. However, to simplify the computation in this procedure, a single preset threshold is used in this process.

• Step 2: Add relief tone pseudo-depth to the object

In REAL, the smallest depth is applied to the central region of the objects judged to be in the foreground, while their perimeter lines are given depths that connect smoothly with the background from the center of the object. Since the REAL depths are considered relative to the viewer, the smallest depth is set to a distance of zero (i.e., no depth). The reason the smallest depth is applied within a fixed range of the center region is to accentuate the jumping-out feeling.

Specifically, the algorithm determines the centroid of each extracted object and uses quadratic curves as expressed by Eq. (1) to represent the change in depth from the centroid to the perimeter.

$$f(x)=\begin{cases}0 & (0\le x\le r_0\cdot l)\\ L\left(\dfrac{x/l-r_0}{1-r_0}\right)^2 & (r_0\cdot l<x\le l)\end{cases} \quad (1)$$

Here, L is the depth of the background ($L\ge 0$), l is the distance from the centroid to the perimeter ($l\ge 0$), r_0 is the ratio of the object allocated to the smallest

depth range ($0 \le r_0 \le 1$), and x is the distance from the centroid ($0 \le x \le l$).

Figure 3 shows the pseudo-depth added to the binary image. Here, $f(x)$ represents the 3-D variation of the depth map, where the protruding regions are shown in lighter shades. Since the regions with greater protrusion are closer to the viewer, the positions in the image with the maximum intensity value have the smallest depth. Taking the distance l from the centroid to the perimeter to be 100%, the object is assigned to the smallest depth from the viewer

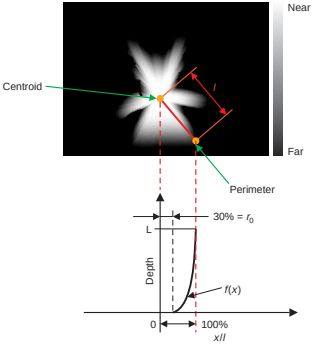


Fig. 3. Adding generated depth by applying quadratic curve.

($f(x)=0$) for a distance determined by r_0 from the centroid (here, 30% of the distance l). Beyond the distance determined by r_0, pseudo-distance is added by applying a quadratic curve so that the distance at the perimeter of the extracted object matches the length of the background. Pseudo-depth is thereby added by performing the abovementioned process for every point on the perimeter.

In REAL it is defined that relief tones must vary smoothly without any sharp changes, but in the results of step 2 there are some parts where the depth is not smooth inside the object due to differences in the distance from the centroid to the perimeter. To remove these sharp depth changes, the depth map is subjected to a smoothing process. In this paper as the filter size we use a relatively large size of at least 0.05% of the image area. This value was obtained by experience from the results of viewing with a DFD display as the size at which a natural 3-D image can be perceived.

3. Comparison of depth generated by REAL with actual depth

Figure 4 compares the depth generated by REAL with the actual depth in typical 3-D images. In REAL the results are close to the actual depth if the depth profile of the subject is spherical, but stray from the actual profile otherwise. For example, in the case of a cylinder the actual depth is only rounded in the horizontal direction, but in REAL, rounding is applied in all directions from the centroid. Also, in the case of two adjacent spheres the actual depth is indented in the central region, whereas in REAL the central region sticks out. Even in cases where the depth pro-

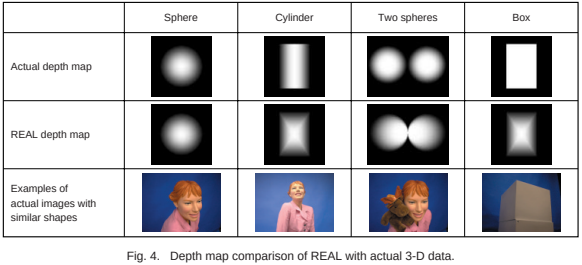


Fig. 4. Depth map comparison of REAL with actual 3-D data.

file is flat, as with a box shape, the results produced by REAL are still rounded in all directions. However, with this degree of discrepancy, when the context of the foreground and background is correct and the depth is continuous and smooth, it is possible to perceive a 3-D impression without any feeling of incongruity by pasting a texture onto this 3-D profile.

4. Characteristics of REAL

The characteristics of REAL are discussed below.

(a) Compact and low cost

The foreground/background separation is a type of distance measurement, but since only rough judgments are made it can be implemented with just a camera and a flash. It is thus easy to make the device compact. Since ordinary compact cameras and flash bulbs perform adequately for this purpose, the equipment can be implemented much more cheaply than conventional 3-D-input devices.

(b) Simple computational processing

REAL uses only simple image processing techniques such as image differencing, thresholding, and quadratic depth interpolation, so it can be fully implemented in real time on commercial PCs. We expect that in the near future even the central processing units (CPUs) incorporated into mobile equipment such as cellphones and personal digital assistants (PDAs) will be able to perform this processing in real time.

(c) Natural image input

REAL can basically operate up to ranges where the light from a flashbulb produces a difference in brightness. Unlike the light sectioning method it does not require a special imaging environment.

(d) Acquisition of texture and depth data with a monocular device

With REAL, a monocular device can acquire texture data and depth data simultaneously. Since it assigns a depth to every pixel in the image (albeit a pseudo-depth), it does not suffer from the occlusion problems that have plagued conventional techniques using triangulation.

(e) Natural 3-D impression

By pasting a texture onto the pseudo-depth of the relief tones, it is possible to obtain a perceptually natural 3-D impression.

5. Subjective evaluation tests

We conducted subjective evaluation tests to examine whether the depth of the relief tones generated by REAL conveys a natural impression to the viewer when presented on a 3-D display.

5.1 Experiments

In the experiments, the test subjects were shown a reference image and a comparative image simultaneously for 15 seconds and were then asked to evaluate the comparative image relative to the reference image on a scale of five levels (very good, good, same, poor, very poor). A blank display period of 5 seconds was inserted before the presentation of the next image pair to prevent the results being affected by the previous evaluation. To make the test subjects aware of the range of fluctuation of 3-D impression, after one practice run, two evaluation runs were performed under the same conditions with the evaluation images presented in random order. A total of nine test subjects were used. An example of the image presentation sequence is shown in **Fig. 5**.

Figure 6 illustrates how the images were presented on the DFD display device in the experiments, and **Fig. 7** shows the images used in the evaluation, which were the typical images shown in Fig. 4. For each evaluation image we prepared a REAL image, a backdrop image, and a planar (2-D) image, which were displayed as shown in Fig. 6. As comparative images we prepared REAL images and backdrop

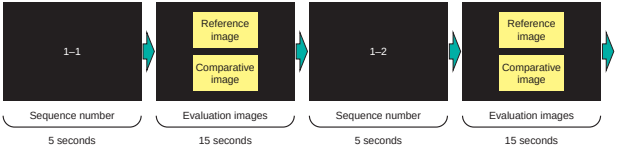
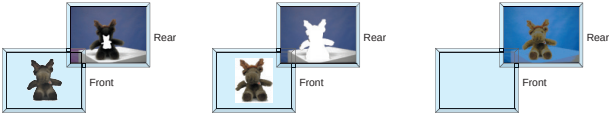


Fig. 5. Evaluation procedure for subjective test.



(a) REAL

Depth-fused 3-D images generated using REAL depth map are displayed in front and rear screens.

(b) Backdrop

The foreground objects are displayed in the front screen. The background objects are displayed in the rear screen like a stage backdrop.

(c) Planar

A 2-D image is displayed in the rear screen alone. The front screen is transparent.

Fig. 6. Types of evaluation images used in subjective test.

Evaluation image 1 | Evaluation image 2 | Evaluation image 3 | Evaluation image 4

Fig. 7. Evaluation images used in the subjective test.

images. The REAL images were 3-D images generated using a REAL depth map. The backdrop images were images in which the foreground object was cut out and displayed in planar form by positioning it in front of its background, which acted like a stage backdrop in a theater. The subjective evaluation experiments were performed by judging criteria in four categories using the planar images as reference images. The evaluated criteria were as follows:

(1) 3-D impression
Does the foreground of the evaluation image stand out compared with the reference image?
(2) Natural impression
Does the evaluation image appear natural compared with the reference image?
(3) Image quality
Does the evaluation image look good compared with the reference image?
(4) Overall
Does the evaluation image seem better than the reference image?

A DFD display device was used for the 3-D display. **Figure 8** shows the setup used in the subjective evaluation tests. The REAL, backdrop, and planar images were scaled to a size of 2 inches and presented at a viewing distance of 300 mm. The image size and

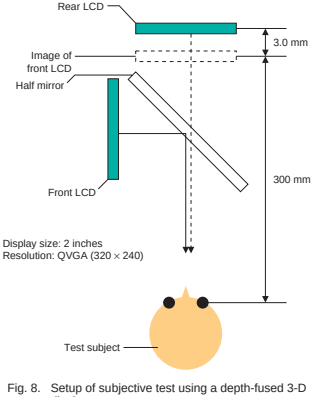


Fig. 8. Setup of subjective test using a depth-fused 3-D display.

viewing distance were chosen to correspond to cellphone usage. The depth representation range of the 3-D display was 3.0 mm.

5.2 Experimental results

Figure 9 shows the results of analyzing the evaluation scores recorded for each of the evaluation criteria. The evaluation scores provided by each test subject were averaged together for each evaluation image when presented as a REAL image, a backdrop image, and a planar image.

Before examining the results relating to each evaluation criterion, we will describe the tendencies of the test subjects. The white circles in Figs. 9(a)–(d) are the theoretical values for the planar images. Since the reference images and planar images were identical, the same evaluation results should in theory be selected for both. However, test subject 4 evaluated this image as "very poor" in one instance. Also, there were many cases in which test subject 9 evaluated this image as "good" or even "very good". A closer look at the evaluation results provided by test subject 4 reveals that this test subject tended to give lower scores than the other test subjects for nearly all the items regardless of whether or not they were presented in planar form. Similarly, test subject 9 tended toward higher scores than the other test subjects. This is assumed to be because different offsets were applied to the evaluation standard by different test subjects even when the same reference values were shown. However, the tendencies in the evaluation of each comparative image (the relative relationship of the evaluation scores) were more or less consistent, so we think that these results can be analyzed by obtaining the average value in this rating scale method.

We examined how the REAL images were perceived by the viewers compared with the conventional image (planar image) for each evaluation criterion. The difference between the scores was confirmed

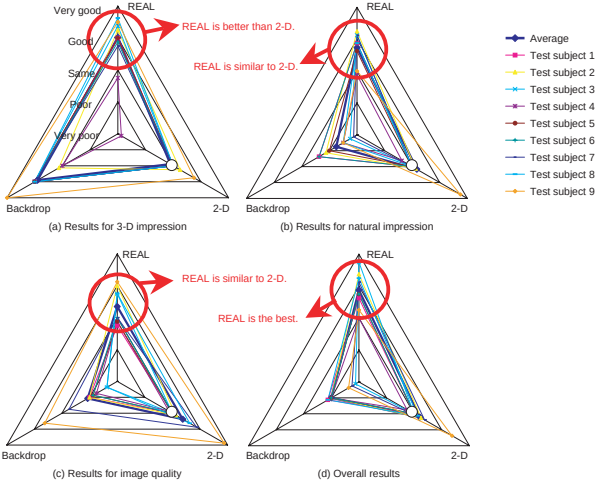


Fig. 9. Results of subjective test.

with a 1% significant difference by performing a *t*-test*.

In the subjective evaluation results for “3-D impression”, the REAL and backdrop images scored better than the planar images. This shows that even with a depth of about 3.0 mm it is possible to perceive a 3-D impression. There was no difference between the REAL images and backdrop images in terms of their 3-D impression.

In the subjective evaluation results for “natural impression”, the planar and REAL images obtained similar scores, while the backdrop images obtained lower scores. This shows that it is possible to perceive a natural 3-D impression similar to that of planar images if pseudo-depth is added in the form of relief tones to only the foreground parts of the image.

In the subjective evaluation results for “image quality”, the planar and REAL images obtained similar scores, while the backdrop images obtained lower scores. Since all the images had the same resolution, we infer that this difference was due to the gaps that appeared in the background of backdrop images due to the cutting away of the foreground parts, thereby causing a reduction in the image quality perceived by the viewers.

In the “overall” subjective evaluation results, the REAL images scored higher than the planar images, while the backdrop images scored lower. The overall evaluation was an evaluation of the viewer’s overall preferences related to the images, and this shows that the REAL images were rated as well as or better than ordinary planar images.

6. Conclusion

We proposed a three-dimensional-input method using a relief-like depth map generation algorithm (REAL). REAL is based on the concept of reproducing a depth map, like the relief of a coin, from 3-D input. It enables a portable 3-D-input system to be achieved using only a monocular camera and a photoflash. REAL is very simple and convenient compared with conventional technology using triangulation. We also gave an overview of the concept and the depth map calculations. Results of a subjective test conducted using 3-D images generated by REAL indicate that observers perceived natural 3-D images on DFD displays. Our experimental results suggest that using REAL for 3-D-input applications in consumer applications will lead to devices that are more compact and have shorter measurement times and less stringent environmental operating criteria, allowing 3-D input to be achieved without accurate distance measurements.

* *t*-test: The two-sample *t*-test is used to determine the difference between the average values of two groups. The meaning of 1% significant difference is that the error of the *t*-test is 1%.

References

[1] “3D Consortium,” http://www.3dc.gr.jp/english/index.html
[2] T. Honda, “About the activity of ‘Consortium of 3-D Image Business Promotion’,” 3D Image Conference 2003, Tokyo, Japan, pp. 205-207, July 2003 (in Japanese).
[3] T. Ando, K. Mashitani, T. Takemoto, M. Higashino, G. Hamagishi, and T. Kobayashi, “A multi-view 3-D display without special glasses,” Technical report of IEICE, EID2002-45, pp. 33-36, 2002 (in Japanese).
[4] I. Harrold, A. M. S. Jacobs, G. J. Woodgate, and D. Ezra, “3D Display Systems Hardware Research at Sharp Laboratories of Europe: an update,” Sharp Technical Journal, No. 74, pp. 24-30, 1999.
[5] A. Miyazawa, “3D imaging technologies for computer-based game machines,” Technical report of IEICE, EID2002-49, pp. 49-52, 2002 (in Japanese).
[6] T. Kotani, “Aiming at pachinko machines and cellphones using 3D displays with semicylindrical lenses with no loss of intensity,” Nikkei Electronics, Jan. 6, 2003, pp. 26-27, 2003 (in Japanese).
[7] S. Suyama, H. Takada, K. Uehira, S. Sakai, and S. Ohtsuka, “A New Method for Protruding Apparent 3-D Images in the DFD (Depth-Fused 3-D) Display,” SID 01 Digest, 54.1, pp. 1300-1303, 2001.
[8] “Special Feature: 3D Data Sensing Technology,” O plus E 20, 11, pp. 1251-1280, 1998 (in Japanese).
[9] K. Sato, “Range sensors,” ITE Journal 54, 2, pp. 157-159, 2000 (in Japanese).
[10] S. Yoshimura, T. Sugiyama, K. Yonemoto, and K. Ueda, “A 48 Kframe/s CMOS Image Sensor for Real-time 3-D Sensing and Motion Detection,” ISSCC 2001 Digest of Technical Papers, San Francisco, California, U. S. A., pp. 94-95, Feb. 2001.
[11] T. Sugiyama, S. Yoshimura, R. Suzuki, and H. Sumi, “A CMOS sensor capable of color video imaging and real-time 3D measurements,” ITE Technical Report 26, 26, pp. 1-16, 2002 (in Japanese).
[12] Y. Yamaguchi, K. Tokai, and T. Iyoda, “Feasibility investigation of mobile 3D image input system based on spatial re-encoding method,” Proceedings of the 2001 Information and Systems Society Conference of IEICE, D-11-67, p. 152, 2001 (in Japanese).
[13] H. Noto and T. Okimura, “Study of 3D-information measurement method using spatial codes,” Proceedings of the 2001 Information and Systems Society Conference of IEICE, D-11-68, p. 153, 2001 (in Japanese).
[14] M. Okutomi and T. Kanade, “A multiple-baseline stereo,” IEEE Trans. Pattern Anal. & Mach. Intell. 15, 4, pp. 353-363, 1993.
[15] S. Kimura, T. Shinpo, E. Kawamura, H. Yamaguchi, and K. Nakano, “A new real-time stereoscopic processing device using spatial filtering,” IEICE PRMU97-207, pp. 1-8, 1998 (in Japanese).
[16] S. Kuwashima, “Color Triclops: The world’s first PC-based full-color stereoscopic vision systems,” Eizojoho Industrial 31, 4, pp. 17-22, 1999 (in Japanese).
[17] S. Shimizu, K. Yamamoto, C. Wang, Y. Sato, H. Tanahashi, and Y. Niwa, “Detection of moving object from mobile stereo omni-directional system (SOS),” Proceedings of 9th Symposium on Sensing via Image Information, Yokohama, Kanagawa , Japan, pp. 531-536, Sep. 2003 (in Japanese).
[18] H. Murata, “Synthesizing techniques of 3D image; 2D-to-3D image conversion technology,” ITE Journal 54, 3, pp. 332-337, 2000 (in Japanese).
[19] “Sharp Corporation SH251iS,” http://www.sharp.co.jp/products/sh251is/ (in Japanese)

[20] "Moneymuseum," http://www.geldauktion.ch/index_english.html
[21] "Kikasete.net questionnaire on camera-equipped cellphones," http://www.kikasete.net/marketer/mk_tw62.php (in Japanese)
[22] H. Eda, "Hit products for the consumer age: Catering for the self-respecting consumer," Nikkei Electronics, Jan. 20, 2003, pp. 59-66, 2003 (in Japanese).
[23] "Commentary: Selection or coexistence? Cellphones vs. digital cameras. Part 1 — User trends: Nikkei BP Consulting Edition, Development centered on non-technical users," Nikkei Mechanical, Sep. 2003 No. 588, pp. 42-43, 2003 (in Japanese).

Hajime Noto
Visual Media Communications Project, NTT Cyber Space Laboratories.
He received the B.E. and M.E. degrees in management engineering from Kansai University, Osaka in 1997 and 1999, respectively. In 1999, he joined NTT Cyber Space Laboratories, Tokyo, where he is engaged in R&D of 3-D-input/output techniques. He is a member of the Institute of Image Information and Television Engineers of Japan (ITE).

Akihiko Hashimoto
Senior Manager, Solution Business Division, NTT Communications.
He received the B.E. and M.E. degrees in electronics engineering from Tokyo Institute of Technology, Tokyo in 1983 and 1985, respectively. In 1985, he joined NTT Human Interface Laboratories, Yokosuka, where he was engaged in research on computer graphics and computer vision systems. He was also engaged in research on 3-D-input systems in NTT Communication Science Laboratories and NTT Cyber Space Laboratories. He is currently working on the IC business for NTT Communications. He is a member of ITE.

Kazuo Kimura
Senior Research Engineer, Visual Media Communications Project, NTT Cyber Space Laboratories.
He received the B.E. and M.E. degrees in electronics engineering from Niigata University, Niigata in 1982 and 1984, respectively. In 1984, he joined NTT Electrical Communications Laboratories, Tokyo, where he worked on active-matrix liquid-crystal displays. Later, he was engaged in the development of Internet facsimile systems and their operation systems in NTT Communications. He is currently engaged in R&D of high-reality systems using 3-D display and image systems. He is a member of ITE.

Kenji Nakazawa
Senior Research Engineer, Visual Media Communications Project, NTT Cyber Space Laboratories.
He received the B.E. and M.E. degrees in electronics in 1981 and 1983, respectively, and the Ph.D. degree in materials science and engineering in 1991, all from Kanazawa University, Ishikawa. He joined NTT Electrical Communications Laboratories, Tokyo in 1983 and was engaged in research on amorphous and polycrystalline Si thin-film materials and TFTs using these films. He was also engaged in R&D of large-screen-size super-high-resolution display systems. He is currently engaged in R&D of 3-D display and image systems. He is a member of the Institute of Electronics, Information and Communication Engineers, ITE, and the Society for Information Display.