Integral three-dimensional television with video system using pixel-offset method

Jun Arai,[1,*] Masahiro Kawakita,[1] Takayuki Yamashita,[1] Hisayuki Sasaki,[1] Masato Miura,[1] Hitoshi Hiura,[1] Makoto Okui,[1] and Fumio Okano[2]

[1]*Science and Technology Research Laboratories, NHK (Japan Broadcasting Corporation), 1-10-11 Kinuta, Setagaya, Tokyo 1578510, Japan*
[2]*NHK Engineering Services, 1-10-11 Kinuta, Setagaya, Tokyo 1578540, Japan*
*arai.j-gy@nhk.or.jp

Abstract: Integral three-dimensional (3D) television based on integral imaging requires huge amounts of information. Previously, we constructed an Integral 3D television using Super Hi-Vision (SHV) technology, with 7680 pixels horizontally and 4320 pixels vertically. We report on improved image quality through the development of video system with an equivalent of 8000 scan lines for use with Integral 3D television. We conducted experiments to evaluate the resolution of 3D images using an experimental setup and were able to show that by using the pixel-offset method we have eliminated aliasing produced by full-resolution SHV video equipment. We confirmed that the application of the pixel-offset method to integral 3D television is effective in increasing the resolution of reconstructed images.

©2013 Optical Society of America

OCIS codes: (110.6880) Three-dimensional image acquisition; (220.4830) Systems design.

References and links

1. M. G. Lippmann, "Épreuves, réversibles donnant la sensation du relief," J. Phys. **4**, 821–825 (1908).
2. Y. Igarashi, H. Murata, and M. Ueda, "3-D display system using a computer generated integral photograph," Jpn. J. Appl. Phys. **17**(9), 1683–1684 (1978).
3. N. Davies, M. McCormick, and L. Yang, "Three-dimensional imaging systems: a new development," Appl. Opt. **27**(21), 4520–4528 (1988).
4. F. Okano, H. Hoshino, J. Arai, and I. Yuyama, "Real-time pickup method for a three-dimensional image based on integral photography," Appl. Opt. **36**(7), 1598–1603 (1997).
5. H. Arimoto and B. Javidi, "Integral three-dimensional imaging with digital reconstruction," Opt. Lett. **26**(3), 157–159 (2001).
6. B. Javidi, I. Moon, and S. Yeom, "Three-dimensional identification of biological microorganism using integral imaging," Opt. Express **14**(25), 12096–12108 (2006).
7. T. Yamashita, M. Kanazawa, K. Oyamada, K. Hamasaki, Y. Shishikui, K. Shogen, K. Arai, M. Sugawara, and K. Mitani, "Progress report on the development of Super-Hi Vision," SMPTE Motion Imaging J. Sept., 77–83 (2010).
8. J. Arai, F. Okano, M. Kawakita, M. Okui, Y. Haino, M. Yoshimura, M. Furuya, and M. Sato, "Integral three-dimensional television using a 33-megapixel imaging system," J. of Disp. Tech. **6**(10), 422–430 (2010).
9. H. E. Ives, "Optical properties of a Lippman lenticulated sheet," J. Opt. Soc. Am. **21**(3), 171–176 (1931).
10. J. Arai, F. Okano, H. Hoshino, and I. Yuyama, "Gradient-index lens-array method based on real-time integral photography for three-dimensional images," Appl. Opt. **37**(11), 2034–2045 (1998).
11. M. Kawakita, H. Sasaki, J. Arai, M. Okui, F. Okano, Y. Haino, M. Yoshimura, and M. Sato, "Projection-type integral 3-D display with distortion compensation," J. SID **18**, 668–677 (2010).
12. Y. Fujita, M. Sugawara, K. Mitani, and T. Saitoh, "A compact, high-performance HDTV camera with four–CCD chips," IEEE Trans. Broadcast **41**(2), 76–82 (1995).
13. M. Sugawara, M. Kanazawa, K. Mitani, H. Shimamoto, T. Yamashita, and F. Okano, "Ultrahigh-Definition Video System with 4000 Scanning Lines," SMPTE Motion Imaging Oct./Nov., 339–346 (2003).
14. M. Kanazawa, K. Hamada, I. Kondoh, F. Okano, Y. Haino, M. Sato, and K. Doi, "An ultrahigh-definition display using the pixel-offset method," J. SID **12**, 93–103 (2004).
15. ISO-12233 standard, "Photography — Electronic still-picture cameras — Resolution measurements," (2000).
16. A. Yasuda, K. Nito, E. Matsui, H. Takanashi, N. Kataoka, and Y. Shirochi, "FLC wobbling for high-resolution projector," J. SID **5**, 299–305 (1997).
17. T. Nagoya, T. Kozakai, T. Suzuki, M. Furuya, and K. Iwasa, "The D-ILA Device for The World's Highest Definition (8K4K) Projection Systems," Proc. IDW 203–306 (2008).
18. H. Hoshino, F. Okano, and I. Yuyama, "Analysis of resolution limitation of integral photography," J. Opt. Soc. Am. A **15**(8), 2059–2065 (1998).

19. J. Arai, H. Hoshino, M. Okui, and F. Okano, "Effects of focusing on the resolution characteristics of integral photography," J. Opt. Soc. Am. A **20**(6), 996–1004 (2003).
20. F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," J. Physiol. **197**(3), 551–566 (1968).

1. Introduction

Integral photography (IP) is a technique proposed by Lippmann for the capture and display of three-dimensional (3D) photographs [1]. After this technique was proposed, remarkable progress has been made in the development of high-resolution imaging and display devices, evolution of fabrication technology for optical elements, and improvement in computer processing performance. Against the background of these developments, there have been many reports on integral imaging related to 3D video systems and video processing techniques based on IP [2–6]. Integral imaging makes it possible to display 3D images in real time without requiring the viewer to use special 3D glasses. Therefore, we are developing integral 3D television systems based on integral imaging.

An integral 3D television requires a large amount of information to display high-quality 3D images. We constructed an integral 3D television based on a full-resolution Super Hi-vision (SHV) system with a resolution of 7680 horizontal pixels times 4320 vertical pixels [7] [8]. With the aim of increasing 3D image quality, we report on the development of a video system with a resolution equivalent to 8000 scan lines and its application to an integral 3D television.

2. Experimental setup

2.1 Configuration

Figure 1 shows the configuration of our integral 3D television system using a camera and projector with a resolution equivalent to 8000 scan lines. The image capture equipment comprises a camera with a resolution equivalent to 8000 scan lines, a lens array, a conversing lens and a depth control lens. When capturing images, the depth control lens is first used to generate a real image of the object. This makes it possible to adjust the depth position of the 3D image when it is displayed. For example, if the real image of the object is generated on the camera side of the lens array, then the 3D image will be generated in front of the lens array. The imaging camera acquires a group of elemental images corresponding to this real image. A problem with this configuration is that when the image capture and display devices both use convex lens arrays, a pseudoscopic image is produced in which the depth direction is inverted [9]. To prevent this pseudoscopic image problem, we configured the lens array using gradient-index lenses [10] or concave lenses. Table 1 lists the specifications of the lens array for image capturing equipment. A conversing lens is placed between the lens array and camera so that the light from the lens array is efficiently directed towards the camera.

In the image display equipment, the group of elemental images captured by the camera is projected onto a diffusion screen using a projector device with an equivalent resolution of 8000 scan lines. A lens array consisting of convex lenses is placed in front of the diffusion screen to generate the 3D image. Table 2 lists the specifications of the lens array for image display equipment. Any distortion in the image projected on the diffusing screen will degrade the reconstructed image produced by the system. Therefore, in this system, degradation of the reconstructed image is prevented by electronically correcting the distortion [11]. The first procedure of the distortion correction process is to measure differences between ideal position and the projected image's position at 11(H) x 9(V) typical points in displayed image. Second, the correction coefficient of each pixel is calculated using polynomials approximating typical point coefficients. Finally, the distortion of the projected image is corrected by the use of the correction coefficient of each pixel. The distance between the diffusion screen and lens array is arranged so as to more or less match the focal length of the convex lenses.
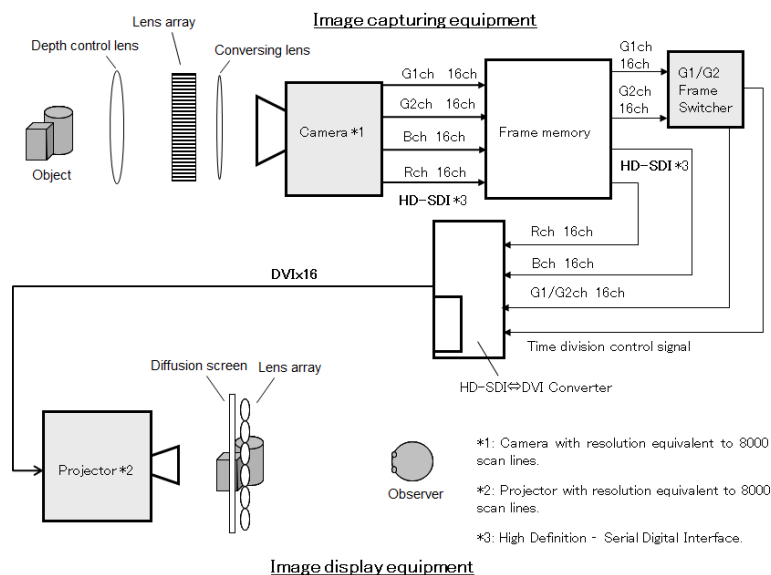
Depth control lens

Lens array

Conversing lens

Image capturing equipment

Object

Camera *1

G1ch 16ch
G2ch 16ch
Bch 16ch
Rch 16ch

HD-SDI *3

Frame memory

G1ch 16ch
G2ch 16ch

G1/G2 Frame Switcher

HD-SDI *3

Rch 16ch
Bch 16ch
G1/G2ch 16ch

DVIx16

Time division control signal

HD-SDI⇔DVI Converter

Diffusion screen   Lens array

Projector *2

Observer

*1: Camera with resolution equivalent to 8000 scan lines.

*2: Projector with resolution equivalent to 8000 scan lines.

*3: High Definition - Serial Digital Interface.

Image display equipment

Fig. 1. Configuration of experimental setup.

Table 1. Specifications of lens array for image capturing equipment

Number of lenses	400(H) x 250(V)
Lens pitch	1.14mm / 1.35mm (Horizontal direction)
Focal length	−2.65mm / −2.745mm
Arrangement	Delta array

Table 2. Specifications of lens array for image display equipment

Number of lenses	400(H) x 250(V)
Lens pitch	1.44mm (Horizontal direction)
Focal length	2.745mm
Arrangement	Delta array

2.2 Imaging camera

There have not been any reports of video systems with a resolution higher than full-resolution SHV. We therefore constructed a camera device with a resolution equivalent to 8000 scan lines by applying the pixel-offset imaging method [12,13] to a 33-megapixel image sensor that we previously developed earlier [7]. Figure 2 shows the configuration and appearance of the camera device, and Table 3 lists its specifications. The camera device uses one CMOS sensor for each of the red and blue signals, and two CMOS sensors for the green signals (G1 and G2 signals). The two green sensors are diagonally offset from each other by half a pixel width to provide a resolution equivalent to 8000 scan lines.

Light entering the camera lens from the object is split by a four-chip color separation prism into two green beams, a red beam (R) and a blue beam (B). These beams are captured by four separate 33-megapixel CMOS sensors. Since green light makes the largest contribution to the luminance component of an image, the luminance signal range of the video system can be effectively expanded with a small number of pixels by using the pixel-offset imaging method [14]. Figure 3 illustrates the sampling structure used for the green signal. The Nyquist frequency for the two sensors in the horizontal and vertical directions becomes double that of one sensor by offsetting their relative spatial relationship diagonally by half a pixel [13].

We conducted tests to evaluate the camera's resolution characteristics. We arranged a resolution test pattern to cover a quarter of the camera's overall field of view in both the

horizontal and vertical directions, as shown in Fig. 4. This resolution test pattern is for use with HDTV systems. Accordingly, the actual resolution values measured in these tests were four times those shown on the test pattern. For example, the figure "1200" printed on the test pattern actually corresponds to a resolution capable of being displayed with 4800 scan lines. Figure 5 shows enlarged views of parts of the images obtained with the camera. Without the pixel-offset method, the limiting resolution corresponds to about 4320 scan lines. The captured G1, G2, and G1/G2 images were interpolated by using the zero-order hold so that each image had the same pixel counts, as shown in Fig. 5. Without the pixel-offset method, aliasing was observed in the pattern corresponding to 4800 scan lines (Fig. 5(a)). However, the pixel-offset method has made it possible to resolve patterns corresponding to 5600 scan lines (Fig. 5(b)).

Figure 6 shows the results of measuring the modulation transfer function (MTF) of the camera including the lens. For these measurements, we used the technique specified in ISO-12233, where the spatial frequency response is obtained from slanted edges in the test chart [15]. The horizontal axis in the figure shows the spatial frequency, and the vertical axis shows the measured MTF. A spatial frequency of 0.5 corresponds to 4320 scan lines. Without the pixel-offset method, the Nyquist frequency is equivalent to 0.5 cycles/pixel. The figure shows that pixel-offset in the horizontal and vertical directions results in a response of approximately 0.1 at a spatial frequency of 0.7 cycles/pixel.
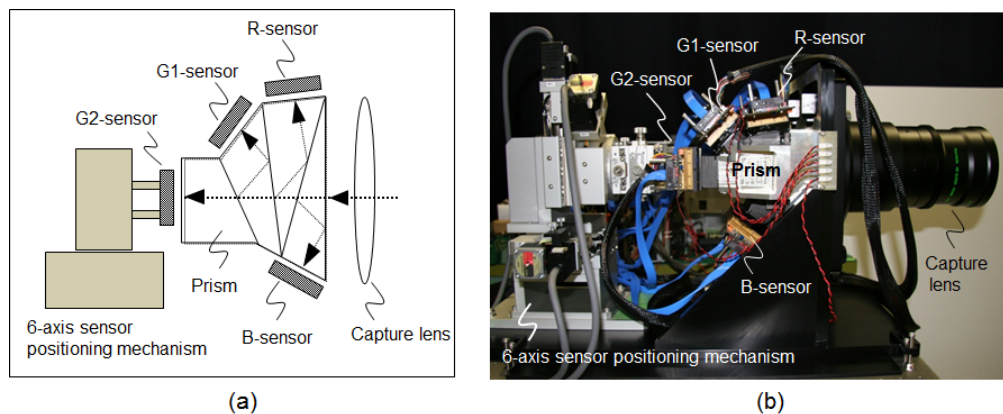

(a)   (b)

Fig. 2. Camera with a resolution equivalent to 8000 scan lines. (a) Configuration. (b) Appearance.

Table 3. Specifications of camera device

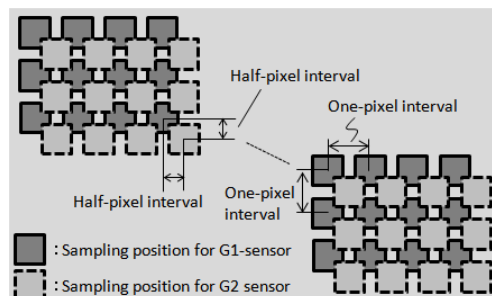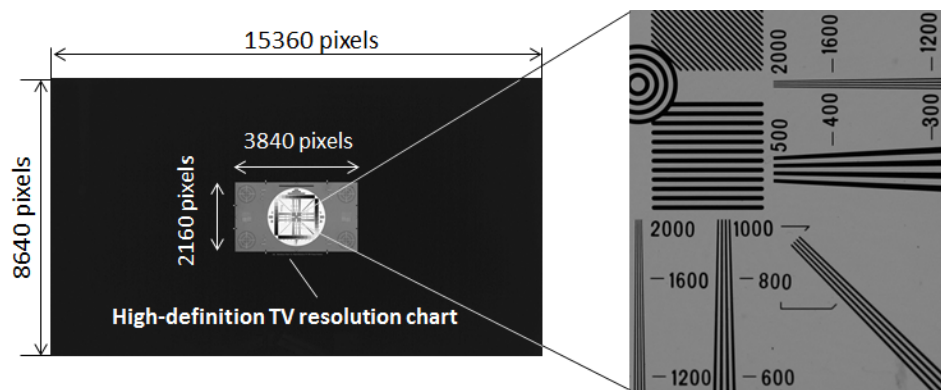| Pixel count | 7680(H) x 4320(V) x 4 |
| | One each for red, blue, green1, and green2. |
| Capture lens | Focal length 61.59mm |

Fig. 3. Sampling structure of green signals.

Fig. 4. Chart used in the tests to evaluate the camera's resolution characteristics.

Fig. 5. Close-up of the chart as captured by the camera. (a) Without pixel-offset method. (b) With pixel-offset method.
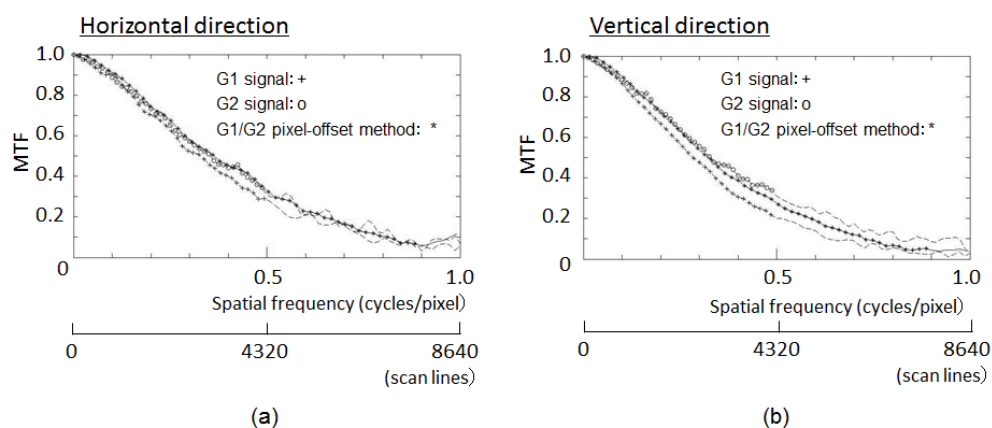

Fig. 6. Results of camera MTF measurements. (a) Horizontal direction. (b) Vertical direction.

2.3 Display projector

The imaging camera device captures the G1 and G2 signals at 60 frames per second with a relative diagonal offset of half a pixel between the CMOS image sensors. A signal switcher switches between the G1 and G2 signals every sixtieth of a second and inputs the resulting signal into the interface converter, as shown in Fig. 1. In the projector, a wobbling element is arranged so as to shift the positions at which the G1 and G2 signals are displayed by half a pixel diagonally every sixtieth of a second. The wobbling element is configured from a polarizing filter and a crystal filter [16]. Due to the movement of the wobbling filter, the green signal is displayed with a resolution equivalent to 8000 scan lines at a rate of 30 frames per second. The red and blue signals are displayed as video with 4320 scan lines at a rate of 60 frames per second. In this case, as with the imaging camera, since the green light makes the largest contribution to the luminance component of the image, it is possible to expand the luminance signal range of the video system with a small number of pixels by using the pixel-offset method [14]. The configuration and appearance of the display projector are shown in Fig. 7. Table 2 lists the specifications of the image display equipment. The LCOS display element has an effective resolution of 7680 (H) × 4320 (V) pixels for each of the red, blue and green signals [17].
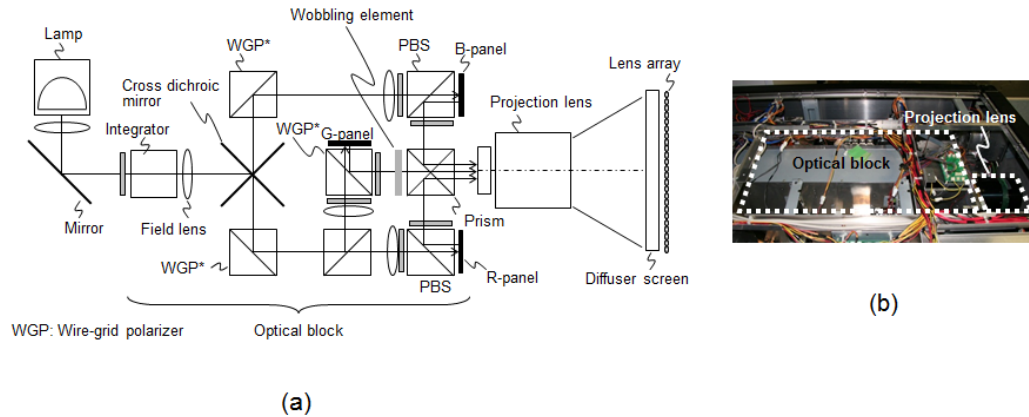

Fig. 7. Projector with a resolution equivalent to 8000 scan lines. (a) Configuration. (b) Appearance.

Tests were conducted to evaluate the resolution characteristics of the projector. We arranged a resolution test pattern to cover an eighth of the overall display angle in both the horizontal and vertical directions, as shown in Fig. 8. This resolution test pattern is designed for use with HDTV systems. Accordingly, the actual resolution values measured in these tests were eighty times those shown on the test pattern. For example, the figure "60" printed on the test pattern actually corresponds to a resolution capable of being displayed with 4800 scan lines. Figure 9 shows enlarged views of the images obtained with a digital still camera. Without the pixel-offset method, the limiting resolution corresponds to about 4320 scan lines. Without the pixel-offset method, aliasing degradation can be seen in the pattern corresponding to 4800 scan lines (Fig. 9(a)). However, the application of the pixel-offset method made it possible to resolve patterns corresponding to 5600 scan lines (Fig. 9(b)). By temporally combining the G1 and G2 signals, the aliasing components are suppressed and the resolution is improved in both the horizontal and vertical directions.
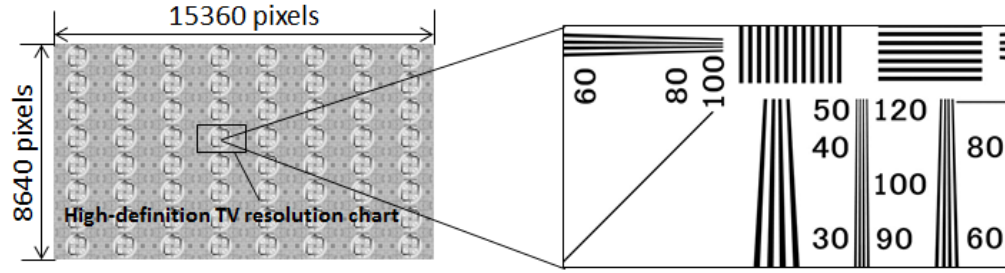
Fig. 8. Chart used in the tests to evaluate the projector's resolution characteristics.

Fig. 9. Close-up of the chart as captured by the projector. (a) Without pixel-offset method. (b) With pixel-offset method.

2.4 Resolution characteristics

In integral imaging, elemental images projected by the lens array are superimposed. The superimposed images generate reconstructed images. In the experimental setup, the maximum spatial frequency of the reconstructed image is limited by the pixel pitch of the elemental images. If p is the pixel pitch of the display element (the pixel pitch on the diffusion screen in the configuration of Fig. 1), the maximum reconstructed spatial frequency can be expressed as follows:

$$\beta = (L - z)|g| / 2p|z| \qquad (1)$$

where β is the maximum reconstructed spatial frequency, g represents the distance from the lens array to the diffusion screen where the elemental images are projected, L is the viewing distance (the distance from the lens array to the viewer), and z is the distance from the lens array to the reconstructed image. Positions to the right of the lens array take positive values, and positions to the left of the lens array take negative values in Fig. 1.

Furthermore, when viewing a reconstructed image produced by the integral imaging technique, the reconstructed image is sampled at the pitch of the elemental lenses. It is therefore necessary to consider the pitch of the elemental lenses that constitute the lens array. The maximum spatial frequency of the reproduced image is restricted by the Nyquist frequency determined by the elemental lens pitch p_L,

$$\beta_n = L / 2p_{\mathrm{L}} \qquad (2)$$

From the above, the upper-limit of the spatial frequency of a reconstructed image produced at an arbitrary depth position can be expressed as follows [18,19]:

$$\gamma = \min\left[\beta_n, \beta\right] \qquad (3)$$

Here, the spatial frequency γ expressed by Eq. (3) is called the "upper-limit spatial frequency."

The dotted line in Fig. 10 shows the relationship between the depth position of the reconstructed image and γ of the reconstructed image in an integral 3D television system corresponding to full-resolution SHV (equivalent to 4320 scan lines). The screen height used in this figure is approximately 312 mm, the viewing distance is set to three times the screen height, and the spatial frequency is expressed in units of cycles per degree (cpd). The spatial frequency of the reconstructed image is largest near the lens array and drops off with increasing distance from the array. The solid line in Fig. 10 shows the corresponding relationship for a video system with an equivalent of 8000 scan lines. In the experimental setup, β_n was 11.34 cpd. When using a full-resolution SHV system, the depth range within which it is possible to generate reconstructed images with β_n is 56 mm. However, this range increases to 106 mm when using a system with an equivalent of 8000 scan lines. Thus, by applying the pixel-offset method, the depth range within which it is possible to generate reconstructed images with the Nyquist frequency can be increased to approximately twice that of a system using full-resolution SHV.

3 Image display experiments

3.1 Resolution

In order to evaluate the resolution of 3D images, we measured the overall MTF of the experimental setup by combining the image capturing equipment and display equipment of an integral 3D television. The procedure for measuring of the overall MTF of the experimental setup is shown in Fig. 11. In step 1, a slanted edge was captured by the image capturing equipment, and the reconstructed image of the slanted edge was generated by the image display equipment. In step2, the reconstructed image of the slanted edge was captured with a digital still camera. The distance from the lens array to the digital still camera was set to three times the screen height. In step 3, from the images captured by the digital still camera, we calculated an MTF reflecting the combined characteristics of the experimental setup and the
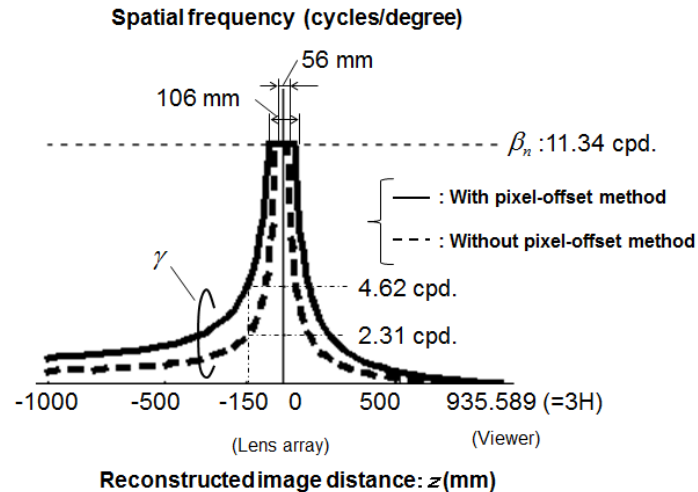
Fig. 10. Theoretical values of the resolution characteristics of reconstructed images.

digital camera. In step 4, we measured an MTF of the digital still camera. Finally, in step 5, the MTF value of the experimental setup itself was obtained by dividing this combined value by the MTF of the digital still camera. When the pixel apertures have a finite width, the high-frequency component of the video signal represented by discrete pixels is degraded. Thus, in this experiment, elemental images were obtained after processing to compensate for the aperture-related degradation of high-frequency components.

The MTF of the experimental setup is shown in Fig. 12, which shows the MTF when the reconstructed image of the slanted edge was reproduced at a distance of -150 mm from the lens array. The horizontal axis shows the spatial frequency of the reconstructed images as seen from the viewer's perspective, and the vertical axis shows the MTF. In this figure, " $+$ " and " o " denote the MTF measured for the G1 and G2 channels, respectively, and " $*$ " denotes the MTF measured from the signal obtained using the pixel-offset method to combine the signals of the G1 and G2 channels. In the reconstructed image generated at a distance of -150 mm from the lens array, the upper-limit spatial frequency obtained without the pixel-offset method is 2.31 cycles/degree. Accordingly, the reconstructed images that contain spatial frequencies higher than 2.31 cycles/degree are affected by aliasing. On the other hand, when the pixel-offset is applied, even though the spatial frequency of the reconstructed image exceeds 2.31 cycles/degree, it remains unaffected by aliasing distortion.

We measured the MTF of the experimental setup when the distance from the lens array to the reconstructed image is set to z = -150, -100, -80, -60, 60, 80, 100, and 150 mm. When the spatial frequency of the reconstructed image is sufficiently low compared to the upper limit spatial frequency γ, relatively high response is gained. We extracted the spatial frequency corresponding to the response of 0.7 from measured MTF values at each depth, and plotted these spatial frequencies in Fig. 13(a). In this case, approximately equivalent response is gained for both the system with and without the pixel-offset method. When the spatial frequency of the reconstructed image is approaching to γ, the response is getting lower. In particular, when the spatial frequency of the reconstructed image is correspond to the one beyond γ, the actual spatial frequency of the reconstructed image is shifted to lower spatial frequency than γ by the aliasing distortion. We extracted the spatial frequency corresponding to the response of 0.05 from measured MTF values at each depth, and plotted these spatial frequencies in Fig. 13(b). In this case, the spatial frequency of the reconstructed image is beyond γ for the system without the pixel-offset method. As a result, the undistorted reconstructed images are not generated. On the other hand, when the pixel-offset method was applied, since the spatial frequency of the reconstructed image is under γ for the system with the pixel-offset method, the undistorted reconstructed images are generated.

The minimum perceptible contrast by the human eye is referred to as the contrast threshold [20]. In [20], experimental results of the contrast threshold of below 0.01 at approximately 5 cycles/degree are reported. The contrast of 0.05 is five times the contrast threshold. Therefore, in this experiment, we chose the response of 0.05 as visually meaningful response. The observer is able to percept adequately the reconstructed images that contain spatial frequencies obtained for MTF value of 0.05. From the above, the result of Fig. 13 (b) shows the effectiveness of applying the pixel-offset method.

3.2 Aliasing

Figure 14(a) shows the reconstructed image of an object consisting of a sinusoidal pattern tilted at an angle. Close-up views of part of the elemental images are shown in Figs. 14(b), 14(c) and 14(d), and the reconstructed images corresponding to these elemental images are shown in Figs. 14(e), 14(f) and 14(g). Figures 14(b) and 14(c) show the elemental images generated from the signals of the G1 and G2 channels, respectively, and Fig. 14(d) shows the elemental images generated from the signal obtained by using the pixel-offset method to combine the G1 and G2 channels. In Fig. 14(d), it can be seen that the elemental image has no aliasing distortion. As a result, the pixel-offset method makes it possible to generate a reconstructed image with no aliasing distortion, as shown in Fig. 14(g).
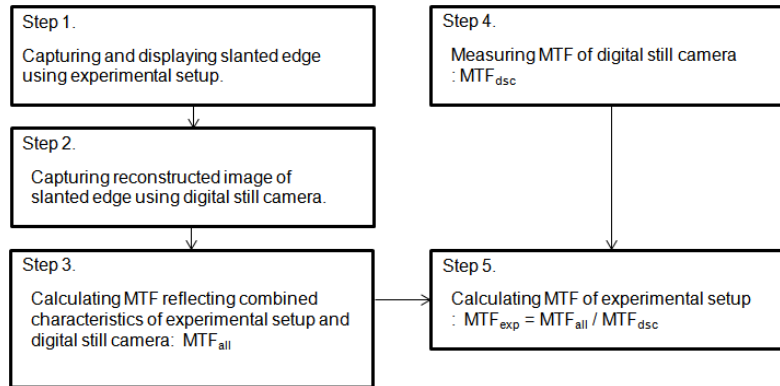
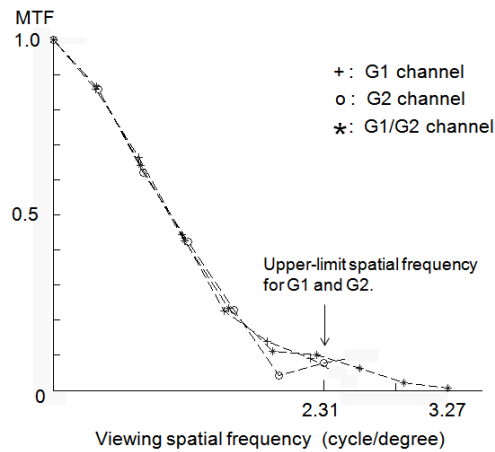Fig. 11. Procedure for measurement of the overall MTF of the experimental setup.

Fig. 12. MTF of the experimental setup. (The reconstructed image of the slanted edge was reproduced at a distance of −150 mm from the lens array.)
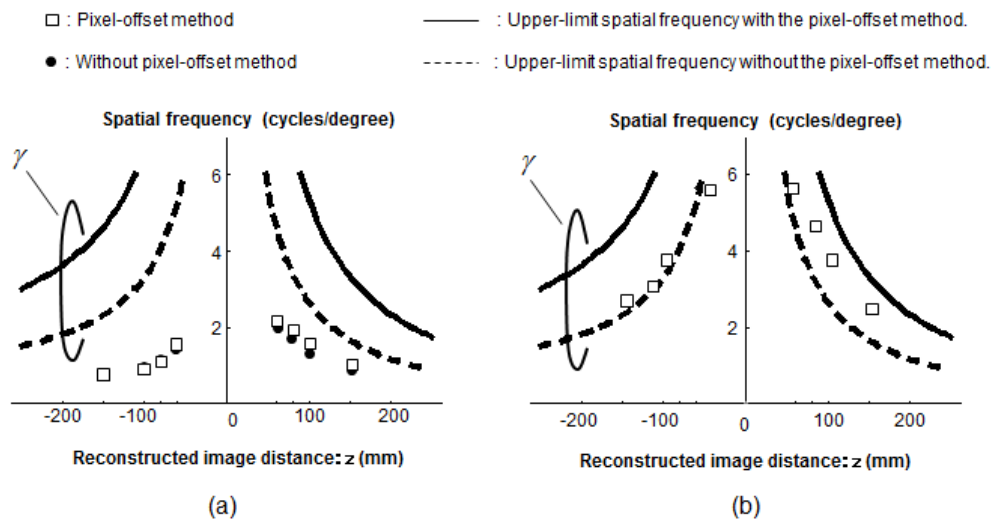

Fig. 13. Relationship between spatial frequency and response in the reconstructed images. (a) Response: 0.7. (b) Response: 0.05.
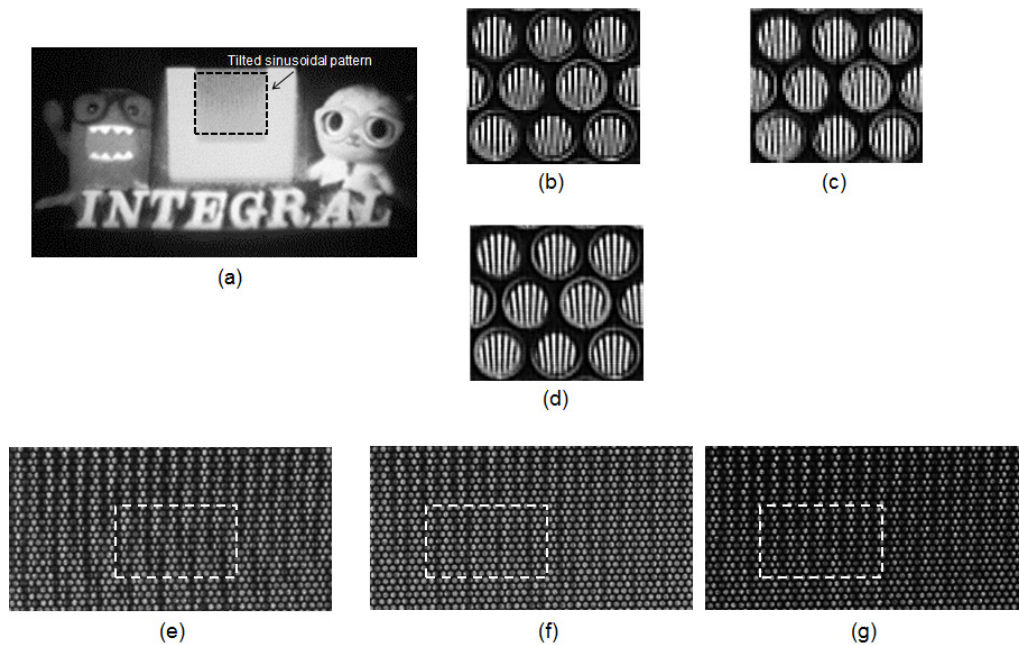
Fig. 14. Reconstructed image and elemental images generated from the experimental setup. (a) Reconstructed image. (b) Close-up of elemental images from the G1 signal. (c) Close-up of elemental images from the G2 signal. (d) Close-up of elemental images obtained by combining the G1 and G2 signals. (e) Close-up of the reconstructed image of the G1 signal. (f) Close-up of the reconstructed image of the G2 signal. (g) Close-up of the reconstructed image obtained by combining the G1 and G2 signals.

3.3 Motion parallax

Figure 15(a) shows the appearance of a reconstructed image, as seen from directly in front of the lens array. With our integral 3D television, the appearance of the picture changes according to the viewer's position from which it is seen. Figures 15(b)–15(e) shows enlarged views of part of a reconstructed image as seen from above, below, to the left, and to the right of the straight-ahead position. As these images show, the reconstructed image adapts to the viewer's position not only in the horizontal direction but also in the vertical direction.
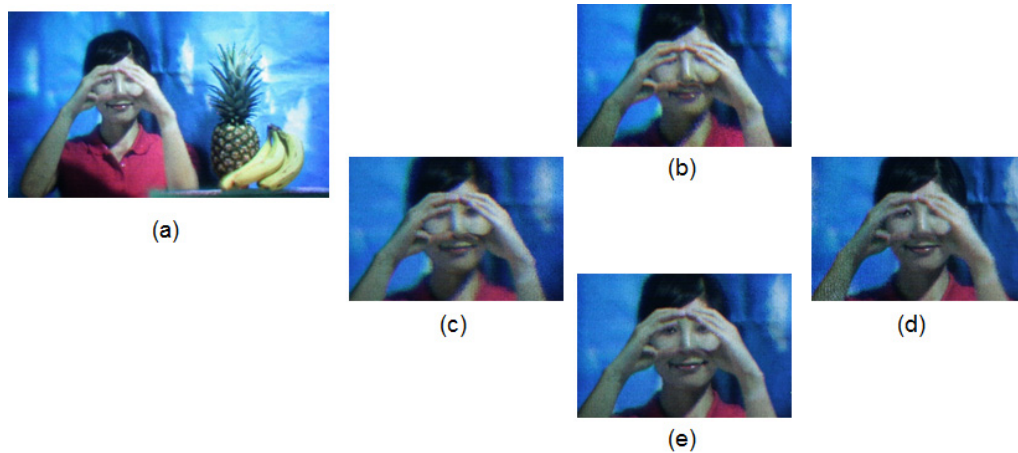
Fig. 15. Changes in the appearance of the image from different viewpoints. (a) View from directly in front. (b) Close-up viewed from upper position. (c) Close-up viewed from left position. (d) Close-up viewed from right position. (e) Close-up viewed from lower position.

4. Conclusion

We developed a video system with a resolution equivalent to 8000 scan lines for an integral 3D television by applying the pixel-offset method. In theory, this method can increase the depth range within which it is possible to reproduce images with the Nyquist frequency to approximately twice that of a system without this method. In order to evaluate the resolution of 3D images, we measured the MTF of an experimental setup that uses the pixel-offset method and confirmed that it is possible to obtain a response of at least 0.05 when the distance from the lens array to the reconstructed image lies from −150 to + 150 mm, even when the spatial frequency of the reconstructed image exceeds the upper-limit spatial frequency of a full-resolution SHV system. The response of 0.05 is adequately high value compared with the contrast threshold. Therefore, the reconstructed image having the response of 0.05 is effective in visual. We also confirmed that when the reconstructed images are viewed, motion parallax is obtained in both the horizontal and vertical directions.

We have thus shown that the application of the pixel-offset method to integral 3D television is effective in increasing the resolution of reconstructed images.

Acknowledgments