Three-dimensional integral television using extremely high-resolution video system with 4,000 scanning lines

Fumio Okano*[a], Masahiro Kawakita, Jun Arai [a], Hisayuki Sasaki [a], Takayuki Yamashita [a], Masahito Sato [b], Koya Suehiro [b], and Yasuyuki Haino [b]
[a] Japan Broadcasting Corporation (NHK); [b] Victor Company of Japan, Limited

ABSTRACT

The integral method enables observers to see 3D images like real objects. It requires extremely high resolution for both capture and display stages. We present an experimental 3D television system based on the integral method using an extremely high-resolution video system. The video system has 4,000 scanning lines using the diagonal offset method for two green channels. The number of elemental lenses in the lens array is 140 (vertical) × 182 (horizontal). The viewing zone angle is wider than 20 degrees in practice. This television system can capture 3D objects and provides full color and full parallax 3D images in real time.

Keywords: integral method, integral television, extremely high-resolution video, full parallax, 3D imaging

1. INTRODUCTION

The integral method is classified as a spatial imaging type. It is considered to be one of the ideal 3D systems because it would enable observers to see a 3D image as though it were a real object. The advantage of this method is that it can produce 3D images using natural light (incoherent light). The integral method was first proposed by G. Lippmann [1] in 1908 based on a photographic technique. Recently, several attempts have been made to obtain higher quality [2-8] and moving 3D images [9-10].

Figure 1 shows the principle of the basic integral method using a single lens array in the capture and display stages. To produce an integral image, a lens array composed of numerous convex elemental lenses is positioned immediately in front of the pickup plate. The integral image is composed of numerous small elemental images that are captured and recorded on the plate. The number of images corresponds to the number of elemental lenses. The integral image is supplied to a transparent display plate. The display plate is placed where the pickup plate had been, and is irradiated from behind by an incoherent light. The light beams passing through the display plate and the lens array retrace the original routes and then converge at the point where the object had been, forming an autostereoscopic image.

The total number of pixels, N_t, on the pickup plate is the product of the number of elemental lenses, N_m, and the number of pixels, N_e, in each elemental image horizontally and vertically:

$$N_t = N_m N_e. \quad (1)$$

The number of elemental lenses determines the upper limit of resolution. The number of pixels in each elemental image affects the resolution of the 3D image away from the lens array. The pickup plate requires N_e times the number of pixels required by conventional television. Thus, extremely high resolution is required for both the capture and display plates. We developed a 3D integral television by applying a video system with 2,000 scanning lines. It provided full color and full parallax 3D images in real time. We have recently developed extremely high-resolution video system with 4,000 scanning lines. Using this, we have improved the characteristics of the 3D television system.

*okano.f-la@nhk.or.jp

Three-Dimensional TV, Video, and Display VI, edited by Bahram Javidi, Fumio Okano, Jung-Young Son,
Proc. of SPIE Vol. 6778, 677805, (2007) · 0277-786X/07/$18 · doi: 10.1117/12.733012

2. IMAGE FORMATION ANALYSIS BY WAVE OPTICS

Although the principle of image formation is described above, a more thorough explanation must include a description in terms of wave optics [11, 12]. Figures 1(a) and (b) focus on around the *m*-th elemental lenses, which collectively constitute an array. A point light source is placed at distance L_s from the lens array and is expressed as a delta function $\delta(x_m - x_{s.m})$, where $x_{s.m}$ represents the object's position. $x_{i.m}$ and $x_{c.m}$ represent the positions in the incident plane of the lens array and pickup plate, respectively. Subscript m indicates the position on the coordinates where the intersection of an incident plane and its own optical axis is the origin for each elemental lens. The following x can be obtained by adding mP_c to x_m, that is, the distance from the origin of the whole array to the optical axis of the elemental lens:

$$x = x_m + mP_c \,. \tag{2}$$

Here, where $x_{s.m}$, $x_{i.m}$ and $x_{c.m}$ are converted to x_s, x_i and x_c, respectively, by adding mP_c in the same way. P_c is the pitch between adjacent elemental lenses in the capture stage. Note that z is not assigned a subscript because the coordinates of each elemental lens match those of the whole array. The origin of the x and z coordinates of the whole array is defined as the point where the optical axis crosses the incident plane of the central elemental lens. To simplify calculations, we use the two-dimensional coordinates (x, z), defined by the x-axis and optical axis z. Real objects in the capture stage can be located in the space with a negative value of z, which is called the real objects area (R.O. area). Real images in the display stage can be located in the space with a positive value of z, which is called the real images area (R.I. area). The following calculations can be applied to the three-dimensional coordinates (x, y, z) defined by the optical axis and a plane that crosses it perpendicularly.

We present the response of the *m*-th elemental lens on the pickup plate shown in Fig. 1(a). First, the wave (electric field) entering the elemental lens of the capture stage is calculated by Fresnel's approximation as

$$\begin{aligned} u_{i.m}(x_{i.m}) &= \frac{1}{j\lambda L_s}\exp\left(-jk\frac{x_{i.m}^2}{2L_s}\right)\int_{object}\delta(x_m - x_{s.m})\exp\left(-jk\frac{x_m{}^2}{2L_s}\right)\exp\left(-jk\frac{x_m x_{i.m}}{L_s}\right)dx_m \\ &= \frac{1}{j\lambda L_s}\exp\left(-jk\frac{x_{i.m}^2}{2L_s}\right)\exp\left(-jk\frac{x_{s.m}^2}{2L_s}\right)e\left(-jk\frac{x_{s.m}x_{i.m}}{L_s}\right), \end{aligned} \tag{3}$$

where $L_s = z_i - z_s$, k is the wave number and equals $2\pi/\lambda$, and λ is the wavelength. The output wave from an elemental lens is a product of Eq. (3) and the phase-shift function of the elemental lens:

$$u_{i.m}(x_{i.m})\exp\left(x_{i.m}^2 \big/ 2f_c\right) \equiv u'_{i.m}(x_{i.m}) \cdot \tag{4}$$

The wave on the pickup plate is obtained by

$$h_{c.m}(x_{c.m}) = \frac{1}{j\lambda g_c}\exp\left(-jk\frac{x_{c.m}^2}{2g_c}\right)\int_{-D_c/2}^{D_c/2} u'_{i.m}(x_{i.m})\exp\left(-jk\frac{x_{i.m}^2}{2g_c}\right)\exp\left(-jk\frac{x_{i.m}x_{c.m}}{g_c}\right)dx_{i.m}\,, \tag{5}$$

where f_c is focal length of the elemental lens, g_c is the gap between the lens array and the pickup plate, and D_c is the width of an elemental lens. The amplitude distribution of this equation is obtained by

$$\left|h_{c.m}(x_{sc.m})\right| = \left|\frac{1}{\lambda^2 g_c L_s}\int_{-D_c/2}^{D_c/2}\exp\left[-jk\left(\frac{1}{L_s}+\frac{1}{g_c}-\frac{1}{f_c}\right)\frac{x_{i.m}^2}{2}\right]\exp(-jkx_{sc.m}x_{i.m})dx_{i.m}\right| = \left|h_{c.m}(-x_{sc.m})\right|\,, \tag{6}$$

$$x_{sc.m} \equiv \frac{x_{s.m}}{L_s} + \frac{x_{c.m}}{g_c}\,. \tag{7}$$

Eq. (6) is an even function and is symmetric for $x_{sc.m} = 0$. The following equation is obtained from Eq. (7):

$$x_{c.m} = -\frac{g_c}{L_s}x_{s.m}\,. \tag{8}$$

This distribution is the same around each point $x_{c.m}$. This point is considered the center of the amplitude distribution for each elemental lens and image point on the pickup plate.

Next, we present the impulse response by the *m*-th elemental lens in the display stage shown in Fig. 1(b). Here, we assume the sale factors of display stage, K_p, K_x, and K_g in terms of elemental lens pitch, elemental image size, and gap between the lens array and the display plate, respectively:

$$P_d = K_p P_c, \qquad (9.1) \qquad x_{d.m} = K_x x_{c.m}, \qquad (9.2) \qquad g_d = K_g g_c. \qquad (9.3)$$

We assume that the point light source $\delta(x_m - x_{d.m})$ is set in the display plate. x_d corresponds to the following position converted from the image point by Eq. (9.2):

$$x_{d.m} \equiv K_x x_{c.m} = -K_x \frac{x_{s.m}}{L_s} g_c = K_x \frac{x_{s.m}}{z_s} g_c, \qquad (10)$$

The input wave of the elemental lens for the display stage is described by

$$u_{d.m}(x_{i.m}) = \frac{1}{j\lambda g_d} \exp\left(-jk\frac{x_{i.m}^2}{2g_d}\right) \int_{elemental\ image} \delta(x_m - x_{d.m}) \exp\left(-jk\frac{x_m^2}{2g_d}\right) \exp\left(-jk\frac{x_m x_{i.m}}{g_d}\right) dx_m \qquad (11)$$

$$= \frac{1}{j\lambda g_d} \exp\left(-jk\frac{x_{i.m}^2}{2g_d}\right) \exp\left(-jk\frac{x_{d.m}^2}{2g_d}\right) \exp\left(-jk\frac{x_{d.m} x_{i.m}}{g_d}\right).$$

The output wave is a product of this equation and the phase-shift function of an elemental lens:

$$u_{d.m}(x_{i.m}) \exp\left(x_{i.m}^2 \big/ 2f_d\right) \equiv u'_{d.m}(x_{i.m}), \qquad (12)$$

where f_d is focal length of the elemental lens, g_d is the gap between the lens array and the pickup plate. Therefore, the output wave in the image space is obtained by

$$h_{d.m}(x_m) = \frac{1}{j\lambda |L_e|} \exp\left(-jk\frac{x_m^2}{2L_e}\right) \int_{-D_d/2}^{D_d/2} u'_{d.m}(x_{i.m}) \exp\left(-jk\frac{x_{i.m}^2}{2L_e}\right) \exp\left(-jk\frac{x_{i.m} x_m}{L_e}\right) dx_{i.m} \qquad (13)$$

$$= \frac{-1}{\lambda^2 |L_e| g_d} \exp\left(-jk\frac{x_m^2}{2L_e}\right) \exp\left(-jk\frac{x_{d.m}^2}{2g_d}\right) \int_{-D_d/2}^{D_d/2} \exp\left(-jk\frac{x_{i.m}^2}{2g_d}\right) \exp\left(\frac{x_{i.m}^2}{2f_d}\right) \exp\left(-jk\frac{x_{i.m}^2}{2L_e}\right) \exp\left(jkK_x \frac{g_p}{g_d}\frac{x_{s.m}}{L_s} x_{i.m}\right) \exp\left(-jk\frac{x_{i.m} x_m}{L_e}\right) dx_{i.m},$$

where, D_d is the width of the elemental lens in the display stage, and $L_e = z - z_i$ is the distance from the array in the image space. The integral operation is performed in the area of an elemental lens. The amplitude of this equation is obtained:

$$|h_{d.m}(x_m)| = \left| \frac{1}{\lambda^2 L_e g} \int_{-D_d/2}^{D_d/2} \exp\left(-jk\frac{1}{2L_f} x_{i.m}^2\right) \exp\left[-jk\left(-K_x \frac{g_p}{g_d}\frac{x_{s.m}}{L_s} + \frac{x_m}{L_e}\right) x_{i.m}\right] dx_{i.m} \right| \qquad (14)$$

$$= \left| \frac{1}{\lambda^2 L_e g} \int_{-D_d/2}^{D_d/2} \exp\left\{-jk\frac{1}{2L_f}\left[x_{i.m} + L_f\left(-K_x \frac{g_p}{g_d}\frac{x_{s.m}}{L_s} + \frac{x_m}{L_e}\right)\right]^2\right\} dx_{i.m} \right|,$$

$$\frac{1}{L_f} \equiv \frac{1}{g_d} + \frac{1}{L_e} - \frac{1}{f_d}. \qquad (15)$$

Equation (14) is the absolute value of the definite integral of a complex Gaussian function and can be modified as

$$|h_{d.m}(x_{se})| = \left| \frac{1}{\lambda^2 L_e g} \int_{-D_d/2}^{D_d/2} \exp\left\{-jk\frac{1}{2L_f}\left[x_{i.m} + L_f\left(\frac{x_s - mP_c}{L_s} + \frac{x - mP_c}{L_e}\right)\right]^2\right\} dx_{i.m} \right| = |h_{e.m}(-x_{se})|, \qquad (16)$$

$$x_{se} \equiv -K_x \frac{g_p}{g_d}\frac{x_s - mP_c}{L_s} + \frac{x - mP_d}{L_e} = -\frac{K_x}{K_g}\frac{x_s - mP_c}{L_s} + \frac{x - mK_p P_c}{L_e}. \qquad (17)$$

$x_{s.m}$ and x are rewritten by the coordinates of the whole array given in the same way as in Eq. (1). The amplitude of the light wave exiting any elemental lens is spread according to Eq. (16). The spread is symmetric for the point $x_{se} = 0$, and we obtain the x coordinate from Eq. (19):

$$x = mP_a\left(K_p - \frac{K_x}{K_g}\frac{L_e}{L_s}\right) + \frac{K_x}{K_g}\frac{L_e}{L_s} x_s. \qquad (18)$$

We can say that this equation shows the light ray from the elemental image.

Here, we assume

$$L_e = \frac{K_p K_g}{K_x} L_s \,, \tag{19}$$

and the following is obtained:

$$x = K_p x_s \,. \tag{20}$$

This x position is independent of the elemental lens number *m-th*, so all waves from the elemental lenses are symmetric for the following point (x_e, z_e), where the waves from all the elemental lenses converge and form an optical image:

$$(x_e, z_e) = \left(K_p x_s, \frac{K_p K_g}{K_x} L_s \right) = \left(K_p x_s, -\frac{K_p K_g}{K_x} z_s \right). \tag{21}$$

When $K_x = K_p = K_g = 1$, the image position is equal to $-z_s$. It is the reversed position of the object for the lens array and the image is pseudoscopic with reversed depth. K_x needs to have a negative value to avoid a pseudoscopic image. Especially when $K_x = -1$ and $K_x = K_p$, the reconstructed image is located at exactly the same position as the object [13, 14].

When $K_x = -K_p$, the depth position z_e is proportional to the scale factor of the gap between the lens array and the display plate, K_g.

The wave at the image point is given by the sum of the squared amplitudes of the waves from each elemental lens, since the light waves emitted from all points $x_{d.m}$ (defined by Eq. (10)) in the display plate are incoherent. Equation (14) is modified at the image plane,

$$\left| h_{d.m}(x_e) \right| = \left| \frac{1}{\lambda^2 L_e g_d} \int_{-D_d/2}^{D_d/2} \exp\left\{ -jk \frac{1}{2L_f} \left[x_{i.m} + \frac{L_f}{L_s} (x_s - x_e) \right]^2 \right\} dx_{i.m} \right| . \tag{22}$$

This equation shows that all amplitudes of the waves from each elemental lens are the same, since they are independent of elemental lens number *m*. The synthesized wave is obtained as

$$\sum \left| h_{d.m}(x_e) \right|^2 = M \cdot \left| h_{d.m}(x_e) \right|^2 , \tag{23}$$

where the summation is performed for all elemental lenses and M is the number of elemental lenses. Therefore, the MTF of the synthesized wave is equal to that of an elemental lens and can be calculated as the Fourier transform of the squared amplitude of the point spread function for an elemental lens.

$$MTF_d(v_s) \equiv F\left(\left| h_{d.m}(x_{e.m}) \right|^2 \right) , \tag{24}$$

where F shows the Fourier transform operation, and v_s (cycles/mm) is the spatial frequency in the image and the object planes.

The MTF of the capture stage is also given by

$$MTF_c(v_c) \equiv F\left(\left| h_{c.m}(x_{c.m}) \right|^2 \right). \tag{25}$$

where v_c (cycles/mm) is the spatial frequency of the elemental image in the capture and display plates. The overall MTF, including the pickup and display stages, is a product of the MTFs of both stages:

$$MTF_t = MTF_c \bullet MTF_d \,, \tag{26}$$

Pixel structure in the capture and display devices affects the MTFs [15]. In the above discussion, we assumed that the pixel pitches are sufficiently fine or that the devices do not have a pixel structure. Therefore, we did not consider the sampling effects of the pixel structure.

3. VISUAL RESOLUTION CHARACTERISTICS

3.1 Visual spatial frequency and examples of MTF

As shown in Figs. 1(a) and (b), spatial frequency α is normalized by the distance [16]. These have the following

relationship between v_s , v_c , and g_c:

$$\alpha = \frac{1}{\tan^{-1}\left[1/\left(v_s|z_s|\right)\right]} \cong v_s|z_s| = v_c g_c \ . \tag{27}$$

The MTF for the capture stage can be expressed as the product of the elemental lens's and the pickup plate's MTFs. The MTF of the display stage can also be expressed as the product of the elemental lenses' and display plate's MTFs. In this section, we assume that the numbers of pixels of these capture and display plates are infinite, meaning the MTF of the elemental lens is the sole factor affecting the resolution. These MTFs are obtained by Eqs. (24) and (25) and are rewritten by α .

The spatial frequency measured from the observer's position, i.e., visual spatial frequency β (cpr), is defined here to clarify the argument. Spatial frequencies α and β have the following relationship [16]:

$$\beta = \alpha\left(L_{OB} - z_s\right)/|z_s| \ , \tag{28}$$

where L_{OB} is the viewing distance between the lens array and the observer. This β is originally defined in the display stage. It can be expanded in the capture stage and considered as a spatial frequency when an object is viewed from the observer's position.

When the observer views the reconstructed image, it is being sampled at the elemental lens, as shown in Fig. 2. The maximum spatial frequency of reconstructed images is limited to the Nyquist frequency. With P_d representing the pitch between elemental lenses, the Nyquist frequency based on the visual spatial frequency can be expressed by

$$\beta_{NL} = L_{OB} / 2P_d \ . \tag{29}$$

The sampling effect is conspicuous when the image is located around the lens array.

Here, we assume a standard observer with 20-20 vision [17]. The observer is usually introduced to evaluate television systems. The resolution threshold corresponds to a resolution of 1.0 minute in a visual angle. The spatial frequency corresponding to the observer's visual acuity, called the visual-limit spatial frequency (VLSF), is given by

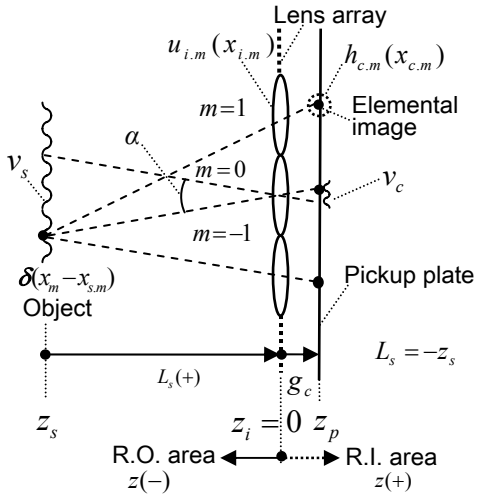
$$\beta_{VL} = \frac{360 \times 60 / 2}{2\pi} = 1720 \ \text{(cycles/rad.: cpr).} \tag{30}$$

The observer cannot perceive an elemental lens alignment with a visual spatial frequency greater than the VLSF. We can set the Nyquist frequency to be greater or less than the VLSF, depending on viewing distance. Here, we set the distance to obtain:

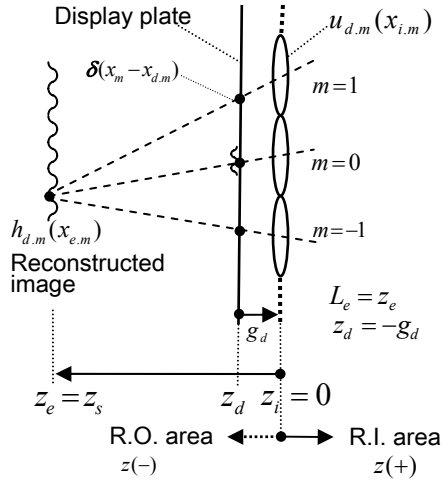
$$\beta_{NL} = \beta_{VL} \tag{31}$$

In addition, actual resolution deteriorates due to an aberration in the elemental lens. However, MTF was calculated using only the influence of diffraction with focus defects because it was assumed that ideal lenses were used as elemental lenses.

MTF can be calculated as the Fourier transform of the squared amplitude of the point spread function. It is equal to calculating the auto-correlation function of the pupil function [18, 19], as is well known. It is assumed that the pupil of each elemental lens is a two-dimensional circle. Figure 3 plots the MTFs in relation to the visual spatial frequencies. The gap between the lens array and the display plate is equal to the focal length of the elemental lens. The diameters of the lenses D_d are 0.25, 0.5, and 1.0 mm. Viewing distances that give the VLSF are 0.86, 1.72, and 3.44 m, respectively. The normalized object distance is 1/4 of the viewing distance. As the graph shows, the smaller the diameter of the elemental lens is, the smaller the MTF is. Figure 4 shows the MTFs, and the normalized object distance is -1/4 of the viewing distance. Other parameters are the same as those in Fig. 3. The graph also shows that the smaller the diameter of the elemental lens is, the smaller the MTF is. The MTFs are improved compared with Fig. 3, because the effect of defocusing is less than the MTFs in Fig. 3. Figure 5 shows the MTF responses in the VLSF. The responses decrease as the object is located at a distance from the lens array. Because of diffraction, this tendency is greater when the radius of the elemental lens is smaller. When the lens is 1.0 mm in diameter, the MTF is not zero even if the object is placed sufficiently far away, at a distance of 100 L_{OB}. This enables high-quality images to be observed.



(a) Capture stage



(b) Display stage

Fig. 1 Principle of basic integral method

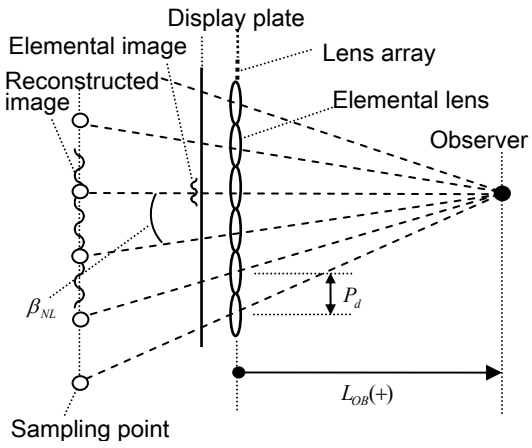


Fig. 2 Sampling by lens array

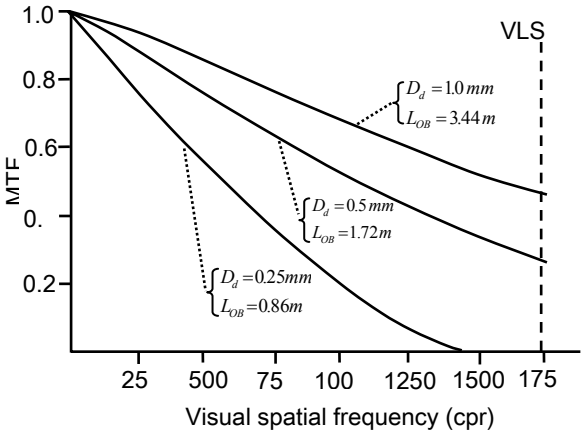


Fig. 3 Relationship between MTFs and visual spatial frequencies
Object is located at a distance of $0.25L_{OB}$ (R.I. area).

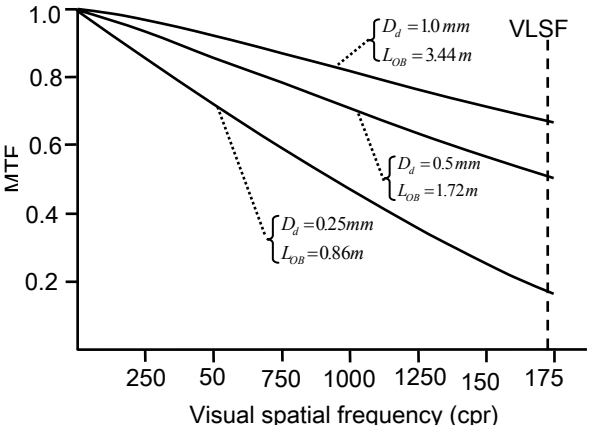


Fig.4 Relationship between MTFs and visual spatial frequencies.
Object is located at a distance of $-0.25L_{OB}$ (R.O. area).

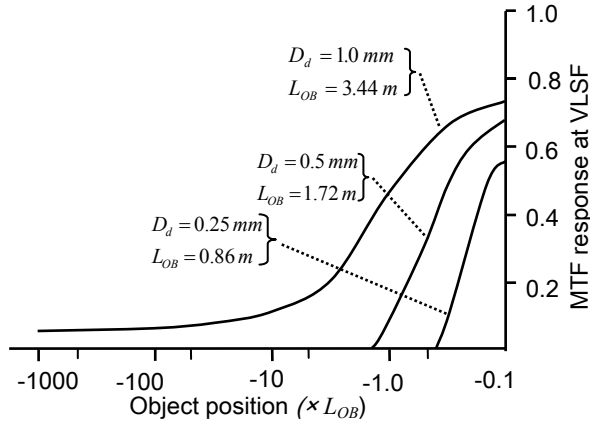


Fig. 5 Relationship between MTF responses at VLSF and object positions

3.2 Effect of pixel pitch

Here, it is assumed the elemental image has a pixel structure, g is the gap between the lens array and the pickup plate or display plate, and the pixel pitch is given by P_{NP}. The elemental image is sampled by the pixels constituting the elemental images. The maximum projected spatial frequency of the object is limited by the Nyquist frequency:

$$\alpha_{NP} \cong \frac{g}{2P_{NP}} \tag{32}$$

Visual spatial frequency β_{NP} according to α_{NP}, is given by Eq. (39):

$$\beta_{NP} = \alpha_{NP} \frac{L_{OB} - z_s}{|z_s|} = \frac{g}{2P_{NP}} \frac{L_{OB} - z_s}{|z_s|} \tag{33}$$

Here, assuming $-z_s$ is infinite, we obtain:

$$\beta_{NP} = \frac{g}{2P_{NP}} \tag{34}$$

The maximum visual spatial frequency β_{MX} which the integral method can produce is given by the following equation.

$$\beta_{MX} = \min(\beta_{DF}, \beta_{NL}, \beta_{VL}, \beta_{NP}) \tag{35}$$

where, β_{DF} is a spatial frequency that gives null response by the diffraction or defocus described in the former section. When the diameter of the elemental lens is more than about 1.0 mm, β_{DF} is larger than β_{VL} of 1720cpr. If the pixel pitch of each elemental image P_{NP} is too large, β_{MX} mainly depends on the Nyquist frequency β_{NP} by the pixel pitch.

The viewing zone angle is given by [19]

$$\phi_{VZ} \cong \frac{P_d}{g_d} \tag{36}$$

A wide viewing zone angle requires small g_d, but small g_d degrades β_{NP}. To compensate for the degradation, P_{NP} needs to be smaller.

4. AN EXPERIMENTAL SYSTEM

4.1 Total system

Figure 6 shows the experimental integral television system. To obtain moving pictures, an image sensor is set on the pickup plate and takes elemental images. For reconstruction, a display plate is placed behind the lens array. The size of the lens array for capturing needs to be large enough to obtain a large parallax, however, it is difficult to develop a large image sensor for moving pictures. In the actual system, a television camera using a taking lens is set to capture whole elemental images formed by the elemental lenses.

A depth control lens, a gradient-index (GRIN) lens array [20,21], a converging lens, and an extremely high-resolution (EHR) camera with 4,000 scanning lines [22] are introduced for capturing. The depth control lens [9] forms the optical image of the object around the lens array, and the GRIN lens array captures the optical image. Many elemental lenses, GRIN lenses, form elemental images near the output plane of the array. An elemental GRIN lens acts as a specific lens that forms an erect image for an object in the distant R.O. area to avoid pseudoscopic 3D images. It works as if K_x is a negative value. The converging lens [9], which is set close to the GRIN lens array, directs the light rays from the elemental GRIN lenses to the EHR camera. The taking lens of the camera is set at the focal length of the converging lens. The converging lens uses light rays efficiently but is not an essential part of the system. In principle, the camera is focused on the elemental images, which are formed around the output plane of the GRIN lens.

Table 1 shows the experimental specifications of the capture stages. Figure 7 shows the two-dimensional alignment of the GRIN lens array used in the experiment. The pitch P_{da} between the adjacent elemental lenses corresponds to 42.5 pixels of the EHR camera. The alignment has a delta structure, which is more efficient than a grid structure. The horizontal pitch P_{dh} is considered $42.5/2$ and the equivalent vertical one P_{dv} is considered $42.5 \times \sqrt{3}/2$.

The video signal of the camera is led directly to a display stage. Table 2 shows the experimental specifications of the display stage. The electronic display setup combines a video projector with diffusion screen and an array of convex lenses. The screen is set as a display plate. The lens array has the same delta structure as the GRIN lens array. The

number of elemental lens is 140 (vertical) × 182 (horizontal). The gap between the lens array and the screen is about the same as the focal length of the elemental convex lens. Comparing Tables 1 and 2, the scale factors are obtained:

$$K_p = 2.32, \qquad K_x = -2.32, \qquad K_g = 2.04$$

The optical image formed by the depth control lens is converted to the following position given by Eq. (21):

$$(x_e, z_e) = (2.32x_s, 2.04L_s). \tag{37}$$

4.2 An extremely high-resolution (EHR) video system with 4,000 scanning lines

We applied an EHR video system with 4,000 scanning lines to the experimental setup. At present, neither image sensor nor display panel with such a huge number of pixels of 4k (vertical) × 8k (horizontal) corresponding to 4,000 scanning lines are available. On the other hand, recently, image sensors and panels with 2k×4k pixels have been developed. Therefore, introducing these is the most feasible way to achieve an EHR system. We equivalently increased the number of pixels in both the horizontal and vertical directions. Both camera and display apply two channels to green (G_1, G_2) using the diagonal-pixel-offset method, red and blue as shown in Fig. 8. Each channel has 2160×3840 pixels. Figure 9 shows the equivalent pixel structure for the scheme that is well known as a Green Bayer Pattern [23]. It takes advantage of human visual characteristics. Specifically, while the human vision features high spatial frequency response with respect to brightness, it has relatively low response with respect to hue and chroma. Because green makes a larger contribution than red and blue to the brightness components, achieving high resolution even in only green can still be effective. It enables an effect roughly equivalent to a video system with 4,000 scanning lines.

The optical system in this camera is composed of a lens and color separation prism, shown in Fig. 10. The prism divides an optical image into G_1, G_2, R, and B on each 2k×4k-pixel image sensor. The four image sensors are attached to the prism with a half-pixel pitch offset so that the image sampling pattern of their pixels is equivalent to that of a single chip color image sensor with 4k×8k pixels, as shown in Fig. 9. A 2160×3840-pixel CMOS image sensor with a 16-mm by 9-mm image area has been applied to each color signal [24]. The vertical limiting resolution of the system can attain more than 3,200 TV lines successfully. We applied a rear projection type display using four LCOS panels [25] in order to obtain a resolution of 4k×8k pixels. Here, the relative positioning of the two LCOS panels for green has a large effect on the resolution characteristics, and these two panels must be accurately offset diagonally by 0.5 pixels. To meet this requirement, we employed two projection units, one for green, where two green panels are installed and the other for red and blue, where red and blue panels are installed. The two green images and R/B images are combined by a half-mirror to obtain one full color image with 4,000 scanning lines. Figure 11 shows the optical system for the green projection unit [26]. Polarizing beam splitters (PBS) with phase-shifting plates are used to separate and combine G_1 and G_2. A compact stepping motor is attached to the G_2 panel to make fine adjustment of the 0.5-pixel offset between these two panels. There are two projection units for dual-G and R/B, the images from which are projected onto the screen. These are accompanied by geometric error. A geometric correction scheme was developed to eliminate this error and can compensate green, red and blue images to match the elemental lens alignment.

4.3 Characteristics of the experimental system

In this experimental television system, the resolution limitation by of the pixel pitch of the elemental image is dominant for distant objects from the lens array because the pixel pitch in the experimental system is too large. The depth control lens shifts the object around the lens array as an optical image. This prevents distant objects from being degraded.

The bandwidth and Nyquist frequency when observer viewed the lens array are shown in Fig. 12. They are normalized by the visual spatial frequency of the pitch P_{da}. There is a difference between the horizontal and vertical directions because the lens array has a delta structure. Figure 13 shows the calculated maximum visual spatial frequency β_{MX} in the horizontal and vertical directions. The viewing distance is assumed to be 2 m. The spatial frequency degrades for objects that are distant from the lens array in the R.O. and R.I. areas.

Figure 14 shows a reconstructed 3D image in a space. The obtained viewing angle with motion parallax is 24.5 degree. Figures 15 (a) and (b) show the 3D images projected on the diffuser set in front of the array. The diffuser, set at a distance of 70mm from the lens array, corresponds to the position of the balloon in Fig.15(a). The balloon is imaged on the diffuser and the letter "k" is blurred. The diffuser is set at a distance of 10mm from the lens array in Fig.15(b). The balloon is defocused and the letter "k" is imaged on the diffuser. These are evidence that the integral imaging reconstructs optical 3D images as holography does.

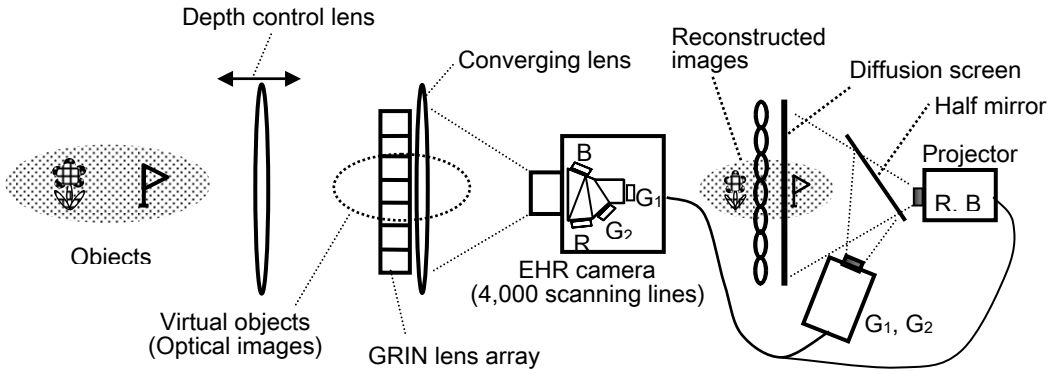


Fig. 6 An experimental integral television system

Table 1 Specifications of capture stage

Camera	Image sensor	4-CMOS (2160(V) × 3840(H) × 4)
	Number of pixels	4320(V) × 7680(H)*
	Size of image sensor	9.1(V) × 16.1(H) mm^2
	Frame frequency	60 Hz
	Focal length of lens	63 mm
	Color system	Color separation prism (R, G_1, G_2, B)
GRIN lens array	Number of GRIN lenses	140(V) × 182(H)
	Diameter	1.085 mm
	Pitch between adjacent lenses	1.140 mm
	Lens length	20.25 mm
	Focal length	-2.65 mm
	Structure	Delta (Line offset)
Elemental image	Diameter	1.085 mm (40.5 pixels*)
	Pitch	1.140 mm (42.5 pixels*)
Depth control lens	Focal length	225 mm
Converging lens	Focal length	650 mm

*: Equivalent pixels using the dual-green diagonal-pixel-offset method

Table 2 Specifications of display stage

Projector	LCD panel	1.7 inch 4-LCOS (2160(V) × 3840(H) × 4)
	Number of active pixels	4320(V) × 7680(H)*
	Frame frequency	60 Hz
	Color system	2 projectors (PBS: G_1/G_2, R/B)
	Projection distance	1,080 mm
	Projection size	21.6 inches
Lens array	Elemental lens	Convex lens
	Number of lenses	140 (V) × 182 (H)
	Diameter	2.64 mm
	Pitch between adjacent lenses	2.64 mm
	Focal length	5.41 mm
	Structure	Delta (Line offset)
Elemental image	Diameter	2.51 mm (40.5 pixels*)
	Pitch	2.64 mm (42.5 pixels*)
Viewing area	Angle	24.5 degrees (measured value)

*: Equivalent pixels using the dual-green diagonal-pixel-offset method

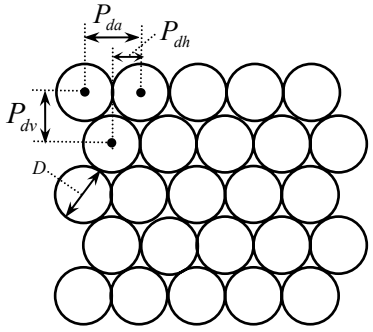


Fig. 7 Elemental lenses alignment of lens array

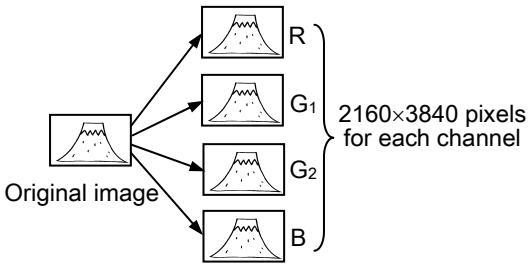


Fig. 8 2G/R/B method

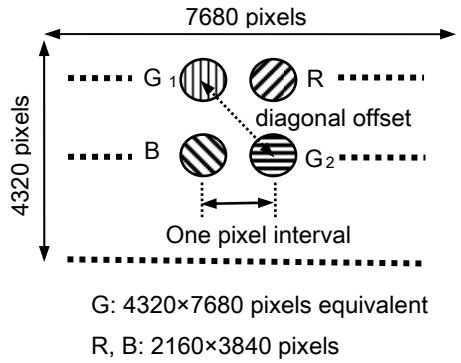


Fig. 9 Pixel alignment on 2G/R/B method in the EHR video system with 4,000 scanning lines

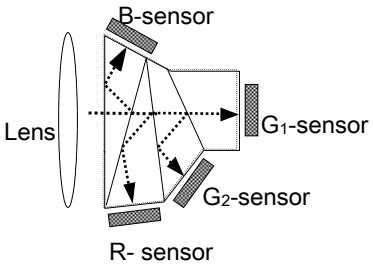


Fig. 10 Color separation prism for 4-image sensors

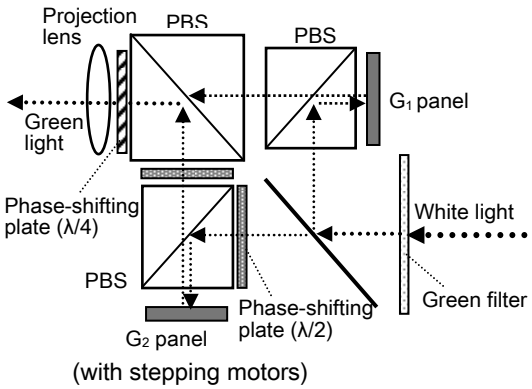


Fig. 11 Optical structure of G projection

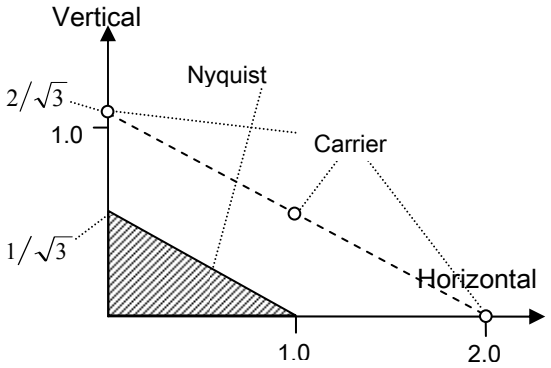


Fig.12 Bandwidth and Nyquist frequency of the lens array

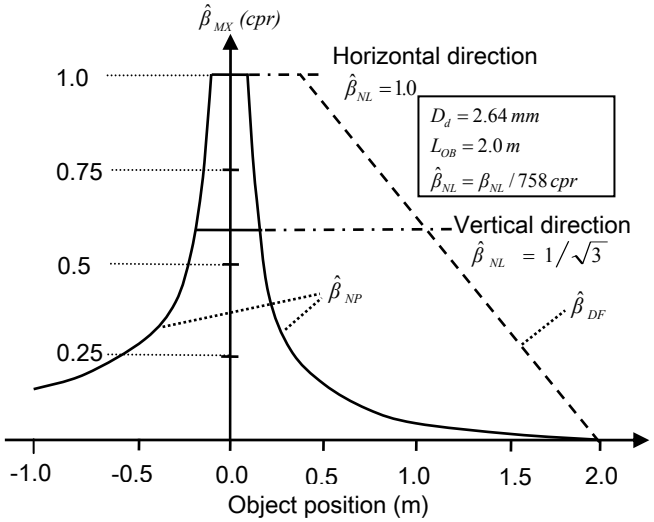


Fig.13 Maximum visual spatial frequency of reconstructed images

Fig.14 Reconstructed 3D image in a space

(a) Projected image on the diffuser set at a distance of 70mm from the lens array

(b) Projected image on the diffuser set at a distance of 10mm from the lens array

Fig.15 Projected images on a diffuser

5. CONCLUSIONS

We described the image formation and resolution for the integral method in the first half of this paper. Diameter of the elemental lens, pitch between the adjacent elemental lenses, human vision, and pixel pitch of the elemental lens all affect the total resolution:

- Small diameter of the elemental lens degrades the resolution of the image located at a distance, because diffraction affects the resolution.
- On the other hand, large diameter of the elemental lens brings also the pitch between the adjacent elemental lenses to be large, and degrades the resolution around lens array.
- The viewing distance is desired to be set so that the diameter of the elemental lens is not perceived.
- Small pixel pitch is required to obtain sufficient resolution with the wide viewing zone angle for the image located at a distance.

Furthermore, the number of elemental lenses would be match the number of pixels in conventional television (for example, 480(V)×640(H) pixels) to obtain proper display size. These factors require the system equips extremely high resolution. Therefore, we introduced an EHR video system with 4,000 scanning lines to an experimental integral television system. The experimental setup produces full-color and full-parallax 3D images in real-time, however, this setup has not practical and not meet the factors yet. A larger number of pixels are required for the capture and display stages to produce higher quality 3D images for television. Although this problem must be overcome, our experimental setup has progressed one step forward in practical terms.

ACKNOWLEDGEMENT

Part of this work has been supported (2006-2010) by National Institute of Information & Communications Technology (NICT).

REFERENCES

1. G. Lippmann, Comptes-Rendus, 146, pp. 446-451 (1908).
2. A. P. Sokolov, "Autostereoscopy and integral photography by Professor Lippmann's method," Izd. MGU, Moscow State University Press (1911)
3. H. E. Ives, "Optical properties of a Lippmann lenticulated sheet," J. Opt. Soc. Am., 21, pp. 171-176 (1931)
4. R. L. DeMontebello, "Wide-angle integral photography – The integral system," in Proc. of SPIE Annu. Tech. Conf., San Diego, Seminar 10, No. 120-08, Tech. Digest, pp. 73-91 (1977)
5. H. E. Ives, "Optical device," U.S. Patent 2 174 003, Sept. 26 (1939)
6. J. Hamasaki, M. Okada, and S. Utsunomiya, "Lens-plate 3D camera using orthoscopic-pseudoscopic image conversion optic," Monthly Journal of Institute of Industrial Science, University of Tokyo, 40(3) pp. 127-136 (1988)
7. N. Davies, M. McCormick, and M. Brewin "Design and analysis of an image transfer system using microlens arrays", Opt. Eng., 33(11), 3624–3633 (1994).
8. J. S. Sang and B. Javidi, "Depth and size control of three-dimensional images in projection integral imaging," Opt. Exp., 12(6), pp. 3778-3790 (2004)
9. 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), pp. 1598-1603 (1997)
10. F. Okano, J. Arai, H. Hoshino, and I. Yuyama: "Three-dimensional video system based on integral photography," Opt. Eng., 38(6), pp. 1072-1077 (1999)
11. F. Okano, J. Arai, and M. Kawakita, "Wave optical analysis of integral method for three-dimensional images," Opt Lett., 2007, 32 (4), 364-366 (2007)
12. F. Okano, J. Arai, and M. Okui, "Visual resolution characteristics of an afocal array optical system for three-dimensional images, Opt. Eng., 46(02), pp.023201 (2007)
13. F. Okano and J. Arai: "Optical shifter for a three-dimensional image by use of a gradient-index lens array," Appl. Opt., 40(20), pp. 4140-4147 (2002).
14. F. Okano, J. Arai, and M. Okui, "Amplified optical window for three-dimensional images," Opt Lett., 31(12) ,1842-1844 (2006)
15. A. Stern and B. Javidi, "Shannon number and information capacity of three-dimensional integral imaging," J. Opt. Soc. Am. A, 21, 1602-1612 (2004).
16. H. Hoshino, F. Okano, H. Isono, and I. Yuyama, "An analysis of resolution limitation of integral photography," J. Opt. Soc. Am. A, 15(8), pp. 2059-2065 (1998)
17. "Relative quality requirements of television broadcast systems," Rec. ITU-R BT.1127
18. H. H. Hopkins, "The Frequency Response of a Defocused Optical System," Proc. Roy. Soc., A231, pp. 91-103 (1955)
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), pp. 996-1004 (2003)
20. F. Okano, J. Arai, and H. Hoshino, "Stereoscopic image pickup device and stereoscopic display device," Japan Patent Appl. No. 08-307763 (1996)
21. 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), pp. 2034-2045, 1998.
22. H. Shimamoto, et al., "An 8k×4k Ultrahigh-Definition Color Video Camera with 8M-Pixel CMOS Imager", SMPTE Motion Imaging J. July /August (2005).
23. B.E. Bayer: Color Imaging Array, US Patent 3971065 (1976)
24. I. Takayanagi, et al., "A 1 1/4 inch 8.3M pixel digital output CMOS APS for UDTV application", Proceedings of ISSCC2003, pp. 216-217 (2003)
25. W.P. Bleha, et al., "D-ILA technology for high resolution projection displays", SPIE AeroSense-03, pp. 5080-29 (2004)
26. M. Kanazawa, et al., "An Ultrahigh-Definition Display using the pixel offset method", J. of the SID 12(1), 93-103 (2004).