

# 3-D computer graphics based on integral photography

T. Naemura, T. Yoshida and H. Harashima

*Dept. of Inform. & Commun. Eng., The Univ. of Tokyo,  
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

*naemura@hc.t.u-tokyo.ac.jp*

<http://www.hc.t.u-tokyo.ac.jp/~naemura>

**Abstract:** Integral photography (IP), which is one of the ideal 3-D photographic technologies, can be regarded as a method of capturing and displaying light rays passing through a plane. The NHK Science and Technical Research Laboratories have developed a real-time IP system using an HDTV camera and an optical fiber array. In this paper, the authors propose a method of synthesizing arbitrary views from IP images captured by the HDTV camera. This is a kind of image-based rendering system, founded on the 4-D data space representation of light rays. Experimental results show the potential to improve the quality of images rendered by computer graphics techniques.

© 2001 Optical Society of America

**OCIS codes:** (100.6890) Three-dimensional image processing, (110.4190) Multiple imaging, (999.9999) Integral photography, (999.9999) Image-based rendering, (999.9999) Light fields

---

## References and links

1. S. E. Chen, "QuickTime VR: an image-based approach to virtual environment navigation," in SIGGRAPH'95, (Association for Computing Machinery, Los Angeles, 1995), 29-38, <http://www.acm.org/pubs/articles/proceedings/graph/218380/p29-chen/p29-chen.pdf>.
2. L. McMillan and G. Bishop, "Plenoptic modeling: an image-based rendering system," in SIGGRAPH'95, (Association for Computing Machinery, Los Angeles, 1995), 39-46, <http://www.acm.org/pubs/articles/proceedings/graph/218380/p39-mcmillan/p39-mcmillan.pdf>.
3. T. Naemura, M. Kaneko and H. Harashima, "Orthographic approach to representing 3-D images and interpolating light rays for 3-D image communication and virtual environment," *Signal Processing: Image Communication*, **14**, 21-37 (1998), <http://www.elsevier.nl/gej-ng/10/22/18/35/17/18/article.pdf>.
4. T. Naemura and H. Harashima, "Ray-based approach to the integrated 3D visual communication," in *Three-Dimensional Video and Display: Devices and Systems*, B. Javidi and F. Okano, eds., *Proc. SPIE* **CR76**, (to be published).
5. 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**, 1598-1603 (1997), <http://www.opticsinfobase.org/ViewMedia.cfm?id=42907&seq=0>.
6. 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**, 2034-2045 (1998), <http://www.opticsinfobase.org/ViewMedia.cfm?id=43189&seq=0>.
7. H. Hoshino, F. Okano, H. Isono and I. Yuyama, "Analysis of resolution limitation of integral photography," *J. Opt. Soc. Am. A*, **15**, 2058-2065 (1998), <http://www.opticsinfobase.org/ViewMedia.cfm?id=1567&seq=0>.
8. T. Yoshida, T. Naemura and H. Harashima, "3-D computer graphics based on integral photography," in *3-D Image Conference 2000*, (Tokyo, 2000), 39-42, in Japanese <http://www.3d-conf.org/00/abstracts/1-9.html#English>.
9. T. Yanagisawa, T. Naemura, M. Kaneko and H. Harashima, "Handling of 3-dimensional objects in ray space," in *1995 Information and System Society Conference*, (Institute of Electronics, Information and Communication Engineers, Tokyo, 1995), D-169, in Japanese.

10. M. Levoy and P. Hanrahan, "Light field rendering," in SIGGRAPH'96, (Association for Computing Machinery, New Orleans, 1996), 31-42, <http://www.acm.org/pubs/articles/proceedings/graph/237170/p31-levoy/p31-levoy.pdf>.
  11. S. Gortler, R. Grzeszczuk, R. Szeliski and M. Cohen, "The Lumigraph," in SIGGRAPH'96, (Association for Computing Machinery, New Orleans, 1996), 43-54, <http://www.acm.org/pubs/articles/proceedings/graph/237170/p43-gortler/p43-gortler.pdf>.
  12. T. Naemura and H. Harashima, "Real-time video-based rendering for augmented spatial communication," in Visual Communication and Image Processing '99, K. Aizawa, R. Stevenson and Y. Zhang, eds., Proc. SPIE **3653**, 620-631 (1999), <http://bookstore.spie.org/cgi-bin/abstract.pl?bibcode=1998SPIE%2e3653%2e%2e620N&page=1&q=spie>.
- 

## 1 Introduction

In the field of computer graphics (CG), the use of real images has been attracting attention as a method of attaining image synthesis that results in a more photo-realistic quality. This field of technology is called image-based rendering [1, 2]. The authors have attempted to solve this problem, by applying the approach of 3-D display technology, in which light rays are the most primitive elements of visual cues [3, 4].

The research reported in this paper focuses on integral photography (IP), which is a 3-D imaging technique. While conventional IP technology is a photographic technique, the NHK (Japan Broadcasting Corporation) Science and Technical Research Laboratories have demonstrated its effectiveness as a media technology using a high definition television (HDTV) camera [5, 6, 7]. Therefore, the authors have proposed and implemented a graphics system that utilizes IP images from an HDTV camera as input, and interactively synthesizes and draws images from various perspectives [8]. This paper reports this system, and investigates its effectiveness. We believe that this research extends the applicability of the optical system for IP, and merges technological field of optical imaging with computer graphics systems.

## 2 Ray-based rendering and IP

### 2.1 Spatial rendering with light-ray data

Holograms and other such 3-D display techniques can be thought of as devices for recording and reproducing the light rays that pass through a display plane. The light rays that pass through a plane in a 3-D space are stored in a four-dimensional data space  $f(x, y, \theta, \phi)$ , with the position on that plane through which the ray passes  $(x, y)$  and the direction of the ray  $(\theta, \phi)$ . Once we obtain such a four-dimensional data space, we can synthesize the images as seen from an arbitrary perspective by selectively reading out light ray data from the data space [9]. In other words, we can simulate the function of 3-D displays, namely the process of reproducing light rays, by computers. In the field of computer graphics, this kind of data space is called light field [10, 11].

Actually, however, efficiently obtaining this four-dimensional data space (the spatial sampling of light ray data) is a difficult task, so some capturing system contrivance or a light-ray data interpolating technique, etc. is required. Especially for dynamic scenes, light rays must be captured simultaneously. For this problem, the authors have previously proposed and implemented a system that performs real-time processing from input to image synthesis [12]. The method is called video-based rendering (VBR). In the VBR system, sixteen CCD cameras are utilized to capture light rays, and the sixteen video signals are combined into one, which is inputted into a graphics computer. As a result of VBR research, we believe that more number of cameras should be densely aligned in order to improve the quality of synthesized images. For this reason, IP system is utilized in this paper.

## 2.2 The evolution of IP technology

IP makes it possible to move the point of view continuously in any direction, without the need for special glasses at the time of observation. So, it has been referred to as one of the ideal 3-D display techniques. A special feature of this system is that it employs an array of microlenses as an optical system to capture and display 3-D scenes with multiple eyes. The microlens array is a set of small lenses, and pieces together an image from the element images formed by the light rays that pass through the center of each lens. Expressed in a different way, it is a technique for recording and reproducing light rays for each direction  $(\theta, \phi)$  at the position of each lens  $(x, y)$  as an element image. So, we can see that this system is highly compatible with the four-dimensional space theory described above.

At the NHK Science and Technical Research Laboratories, a real-time, full-color IP system has been achieved using a gradient-index lens and an HDTV camera [5, 6, 7]. The configuration of that system is illustrated in Fig. 1.

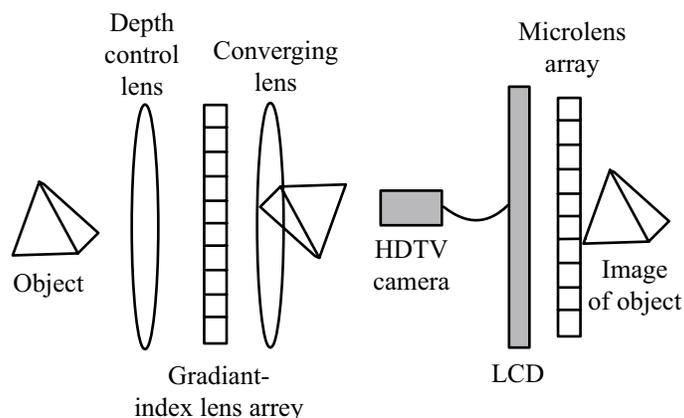


Fig. 1. Diagram of real-time IP system.

The gradient-index lens array (optical fiber bundle) on the imaging side and the lens array that is set on the liquid-crystal display (LCD) panel on the display side correspond to the microlens array of the conventional IP. Use of the gradient-index lens solves the problems of false images and interference (See Table 1 and references [5, 6, 7] for more information). In the HDTV image, the number of inverted real images of the object, obtained as element images, equals the number of lenses (optical fibers). An example of this is shown in Fig. 2. When that HDTV image is presented on the LCD, the view of

Table 1. Equipment specifications [6].

CCD camera	
pixel number	1920(H) $\times$ 1035(V)
pickup device size (mm <sup>2</sup> )	14.0(H) $\times$ 7.9(V)
lens focal length (mm)	55
Lens array	
lens number (off-set structure)	54(H) $\times$ 63(V)
pitch (mm)	1.085
Element image	
pixel number	20(H) $\times$ 20(V)

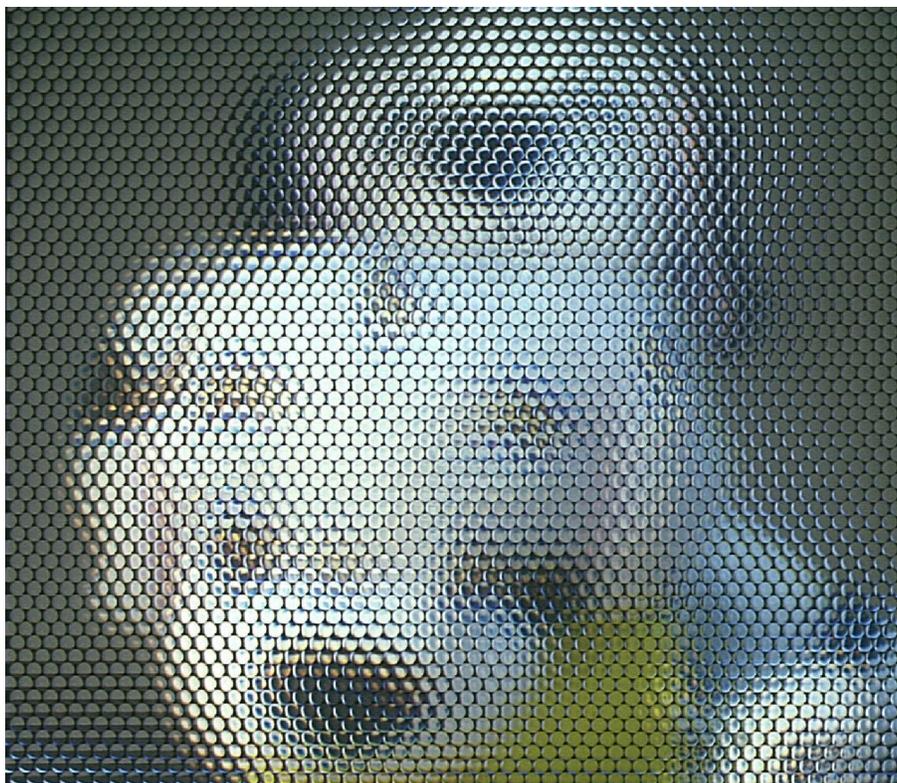


Fig. 2. Captured IP image.

the image changes with the viewer's position as a result of the effect of the microlens array placed on the LCD panel. The system thus functions as a 3-D display device.

Here, the authors investigate a computer graphics system that synthesizes image views interactively according to the viewing position of the viewer from images such as shown in Fig. 2, without using a special optical system on the display side.

### 3 Proposed method

For an input image such as shown in Fig. 2, positional alignment for the center point and size of each element image is done in advance. In this paper, we use a 57-row by 54-column array of element images. A comparison with the VBR [12] described above as a multi-eye simultaneous imaging system is presented in Table 2. We see that the proposed system can be positioned as a system that has a larger number of eyes and a lower element image resolution than VBR. Another point of difference is that, although image synthesis by real-time processing of the D1 video signal is possible with VBR, real-time processing in the input module is not possible with the proposed system, because the input signal is an HDTV signal.

#### 3.1 *Optical system simulation in the reproduction system*

To begin with, we consider the computed synthesis of an image that is equivalent to an image reproduced by the system of the NHK Science and Technical Research Laboratories. For simplicity, we let each lens in the microlens array correspond to one color (i.e., select one pixel from the corresponding element image). We refer to this as **Method 1**.

Table 2. Comparison of VBR and the proposed method.

Method	Eye count	Resolution of the element image
VBR [12]	$4 \times 4$	Approx. $180 \times 120$
IP (Proposed method)	$54 \times 57$	Approx. $20 \times 20$

By assuming that a light ray passing through the center of a lens is not refracted, this pixel selection can be formalized as shown below. That is to say, an  $XYZ$  coordinate system is defined as shown in Fig. 3, with the display represented as  $Z = 0$ , the viewing position as  $(x_0, y_0, z_0)$ , the center position of the element image as  $(x_c, y_c, 0)$  and the  $Z$ -direction distance between the display and microlens array as  $d$ . The position  $(x_e, y_e)$  of the pixel that can be seen through the lens is given by the following equations.

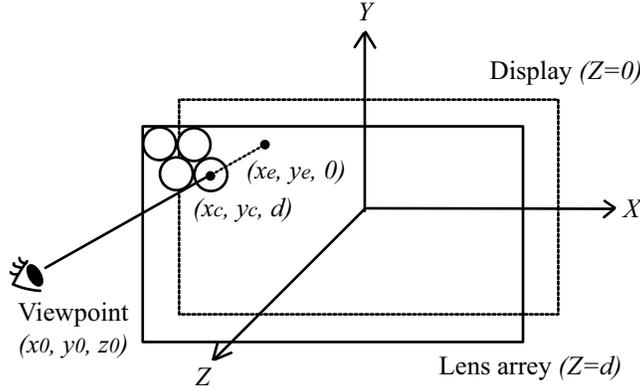


Fig. 3. Coordinate system.

$$x_e = x_c + (x_c - x_0) \frac{d}{z_0 - d} = \frac{z_0 x_c - dx_0}{z_0 - d} \quad (1)$$

$$y_e = y_c + (y_c - y_0) \frac{d}{z_0 - d} = \frac{z_0 y_c - dy_0}{z_0 - d} \quad (2)$$

$$(3)$$

### 3.2 Higher image quality through assumption of depth

Applying the above method makes it possible to reproduce the image by making a one-to-one correspondence between a lens and a color. It is thus possible to synthesize a smoother image by linear interpolation of the colors of three contiguous lens. We call this approach **Method 2**. With linear interpolation, however, we cannot necessarily expect an improvement in image quality.

We therefore implemented a more sophisticated interpolation in which the depth of the object is assumed. We call this **Method 3**. In this method, as illustrated in Fig. 4, the shape of the object is considered to be approximated by an object plane  $S$  ( $Z = z_s$ ) that is parallel to the display plane ( $Z = 0$ ). For simplicity, Fig. 4 shows an  $XZ$  plane aspect of Fig. 3. The following discussions, however, can be also applied to  $YZ$  plane.

In Fig. 4,  $L_m$  and  $L_{m+1}$  denote the  $m$ -th and  $(m+1)$ -th lens, respectively. The center of the lens  $L_m$  is located at  $(x_m, d)$ , and  $L_{m+1}$ ,  $(x_{m+1}, d)$ . A straight line, which pass through viewpoint  $A$  ( $x_0, z_0$ ) and the center of lens  $L_m$  ( $x_m, d$ ), can be denoted as  $AL_m$ . The intersection of the line  $AL_m$  with the object plane  $S$  is denoted as  $Q$ , and that with

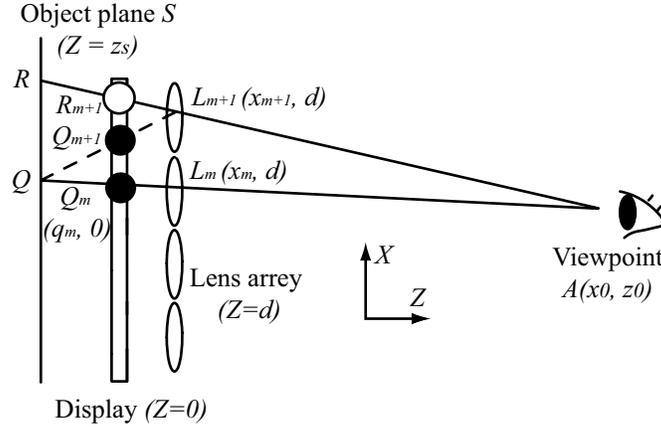


Fig. 4. Principle of interpolation.

the display plane,  $Q_m(q_m, 0)$ . In the same way, we can define an object point  $R$  and a corresponding display point  $R_{m+1}(r_{m+1}, 0)$  by considering a straight line  $AL_{m+1}$ . Through the lens  $L_m$ , we can see the display point  $Q_m$  as the object point  $Q$ . When the object point  $Q$  is also seen through  $L_{m+1}$ , the corresponding display point  $Q_{m+1}(q_{m+1}, 0)$  is the intersection of a straight line  $QL_{m+1}$  with the display plane. Thus, the following relationships can be obtained.

$$\frac{q_m - x_m}{-d} = \frac{x_m - x_0}{d - z_0} \quad (4)$$

$$\frac{q_{m+1} - q_m}{-z_s} = \frac{x_{m+1} - x_m}{d - z_s} \quad (5)$$

If we denote the  $m$ -th element image as  $I_m$  and the pixel at  $(q_m, 0)$  on  $I_m$  as  $I_m(q_m)$ , then, with **Method 1**,  $Q$  and  $R$  are rendered by  $I_m(q_m)$  and  $I_{m+1}(r_{m+1})$ , respectively. With **Method 2**, the color between  $Q$  and  $R$  is approximated by linear interpolation of  $I_m(q_m)$  and  $I_{m+1}(r_{m+1})$ .

In contrast to that, with **Method 3**, the partial image of  $I_{m+1}$ ,  $I_{m+1}(x)$  ( $q_{m+1} \leq x \leq r_{m+1}$ ), can be used as the texture between  $Q$  and  $R$ . Similarly, because  $I_m(x)$  ( $q_m \leq x \leq r_m$ ) can also be used, it becomes possible to synthesize the texture between  $Q$  and  $R$  by proper weighting of the appropriate part of each element image. Real-time texture synthesis can be achieved by application of alpha blending technology.

#### 4 Experimental results

The results of applying **Method 1** with the image shown in Fig. 2 as the input image is shown in Fig. 5. This can be regarded as a simulation result of the optical system. The results for **Method 2** are shown in Fig. 6. Compared to **Method 1**, **Method 2** produces a smoother image, but the overall impression made by the image is not good, so we cannot necessarily say that the image quality has been improved.

The results for **Method 3** are shown in Fig. 7. The left image in the figure is synthesized when the depth parameter  $z_s$  is matched to the dog in the back. For the right image, that parameter is matched to the dog in the front. We created and used a graphical user interface (GUI) that allows the parameter  $z_s$  to be set interactively while the synthesized image is being viewed. We were able to improve the quality of parts of the image, according to the depth of the object. From the nature of the method, the



Fig. 5. Result of **Method 1** ([res1.mov](#); 617 KB). Fig. 6. Result of **Method 2** ([res2.mov](#); 329 KB).

image quality tends to worsen for objects that are not at the assumed depth according to how far from that depth they are.



Fig. 7. Results of **Method 3** ([res3b.mov](#); 291 KB), ([res3f.mov](#); 270 KB).

The results for different input images are shown in Fig. 8 and Fig. 9. Objects that glitter and reflect light in complex ways are not easily drawn with conventional computer graphics technique, that is polygon models, but the effectiveness of the proposed image-based method is demonstrated in Fig. 8. Furthermore, because simultaneous imaging with a multi-eye imaging system is necessary for drawing moving subjects such as people with the image-based method, results that show the effectiveness of using IP images for that case are presented in Fig. 9.

The result images shown in the figures were synthesized in real-time while interactively changing the viewpoint position and the depth parameter of the object,  $z_s$ .

## 5 Conclusions

We have proposed a method for the interactive synthesis of 3-D computer graphics using IP images and have implemented a system based on that method. We investigated methods that apply complex processing by computer to achieve better image quality

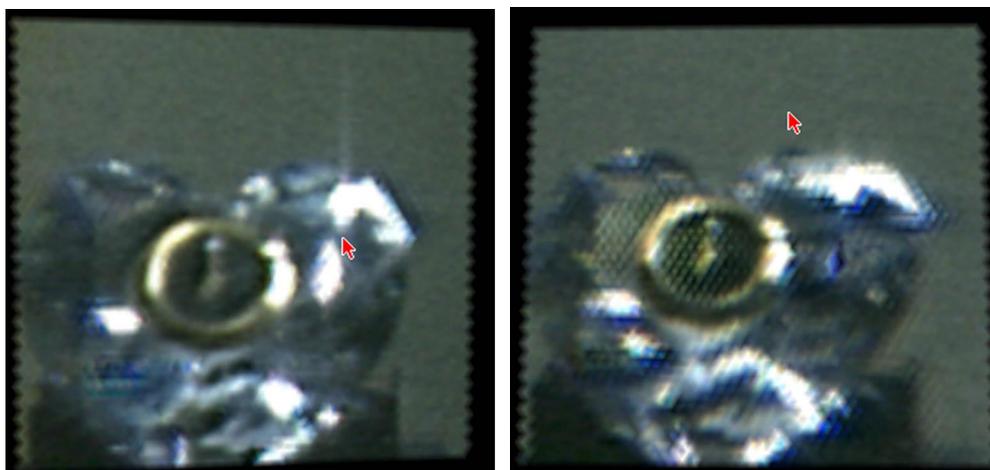


Fig. 8. Result of reflective object ([crystal.mov](#); 360 KB).



Fig. 9. Results of human images ([naemura.mov](#); 178 KB), ([finger.mov](#); 254 KB).

than can be achieved by methods that use special optics to reproduce 3-D images. A method for quantitative measurement of improved image quality remains to be studied. The current system cannot input HDTV video in real time, and so is intended for the processing of static scenes, but in future work, we plan to investigate real-time input of moving images.

### Acknowledgements

We thank Ichiro Yuyama, Fumio Okano, Haruo Hoshino, and Jun Arai of the NHK Science and Technical Research Laboratories for their kind permission to use the IP images employed in this research.