

Progresses in 3D integral imaging with optical processing

Manuel Martínez-Corral¹, Raúl Martínez-Cuenca¹, Genaro Saavedra¹, Héctor Navarro¹, Amparo Pons¹, and Bahram Javidi²

¹Department of Optics. University of Valencia. Calle Doctor Moliner 50, E46100, Burjassot, Spain.

²Electrical and Computer Engineering Dept, University of Connecticut, Storrs, CT 06269-1157

E-mail: manuel.martinez@uv.es

Abstract. Integral imaging is a promising technique for the acquisition and auto-stereoscopic display of 3D scenes with full parallax and without the need of any additional devices like special glasses. First suggested by Lippmann in the beginning of the 20th century, integral imaging is based in the intersection of ray cones emitted by a collection of 2D elemental images which store the 3D information of the scene. This paper is devoted to the study, from the ray optics point of view, of the optical effects and interaction with the observer of integral imaging systems.

1. Introduction

The design and development of techniques for pickup and display of three-dimensional (3D) images has attracted the attention of scientists and engineers in very different disciplines. During 18th century, Smith [1] and Porterfield [2] studied the influence of binocular vision on the disparity between the images perceived and on depth perception as well. In 1838, Wheatstone [3] proposed the first theory on binocular vision and built the first stereoscope. A few years later W. Rollmann designed the first anaglyph [4], based on the concept of complementary colours. However, both the stereoscope and the anaglyph provide only with a single relief perspective of a given scene. Thus, two observers at different locations would perceive the same image.

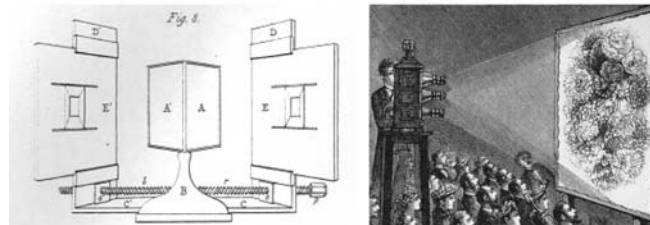


Figure 1. Wheatstone's stereoscope (left) and anaglyphs projection (right) were the first attempts to generate 3D images.

At the beginning of the 20th century, G. Lippmann [5] proposed a novel technique, named as Integral Photography (IP), that allows the generation of true 3D images. This technique, which is

based on the reversibility principle for light rays, is characterized by being auto-stereoscopic, namely, there is no need for the use of any special glasses or any additional device to obtain proper image visualization. A few years later, H. E. Ives [6]-[8] proceeded on Lipmann's research and he proposed the use of lens sheets and parallax barriers instead of fly eye lenses to make easier the building of these systems. During the second half of 20th century, the groups led by Buckhardt, Okoshi and McCornick [9]-[12] continued the study and improvement of the IP designs.

The first attempts for obtaining 3D animated images were severely limited by the low resolution of the digital cameras and video projectors available at the 80's. Besides, the real-time processing of the required amount of information was unreachable for the digital devices available by that time. However, thanks to the advances in microlens-array building techniques and to the development of high resolution digital methods, nowadays the IP concept has been rescued under the name of Integral Imaging (InI) [13]. The initial limitations in resolution seem to be gradually overcome by use of successive generations of digital cameras (CCD and CMOS). The use of computers for the information processing confers to this technique an attraction in applications like Medicine, Telemetry or 3D object recognition. The InI systems are continuously improving their marks in resolution, processing speed and memory, decreasing their cost at the same time.

2. Imaging and resolution in InI

Integral imaging is achieved in two stages, namely, the pickup (or recording) and the display. Pickup step consists on the simultaneous recording of a set of 2D images from a 3D scene. Any individual image is named as elemental image. The whole set of elemental images is called as the integral image of a 3D scene. Each elemental image contains information of a different view of the object. In the display stage the 3D scene is reconstructed by intersection of the ray bundles emanated from each elemental image.

The recording of the set of elemental images can be performed optically in the way described above. But the integral image can be built, as well, by digital methods from a 3D file in which the 3D intensity distribution of the object has been stored. Such digital process is commonly known as virtual recording.

Equivalently, the display stage can be performed optically or by digital methods. In the first case we will refer to the display system as the visualization system. In the digital case we will name the system as the reconstruction one. In any case, the systems used in both stages have to be similar to each other to allow the 3D image to be the closest to the original object. In fact, a break in the symmetry between the geometry in both stages generates distorted 3D images [14].

2.1. Pickup systems

2.1.1. Pinhole array

To analyze the optical pickup process, let us start by considering the scheme shown in Figure 2, in which an array of pinholes, with pitch p , is used. The 3D scene is projected through the pinholes onto the matrix sensor to generate an array of 2D microimages of the 3D scene. The microimages are registered in a matrix sensor located at a distance g from the pinhole array. This distance is referred usually as the *gap* of the InI system. Note that, if only these two elements were used, the microimages would be recorded with no limitation in extension, so that the ones generated by neighbouring pinholes would overlap. To avoid this effect, which is known as overlapping effect, a set of opaque barriers should be used to divide the sensor plane into a set of elementary cells [15], each centred at the corresponding pinhole. The part of a microimage which falls inside of the elemental cell is called as elemental image. If the set the origin for the Cartesian coordinates systems just at the central pinhole, then the positions of the other pinholes are

$$\mathbf{x}_{p,m} \equiv (x_{p,m}, z_{p,m}) = (mp, 0) , \quad (1.)$$

where the integer m is the pinhole index.

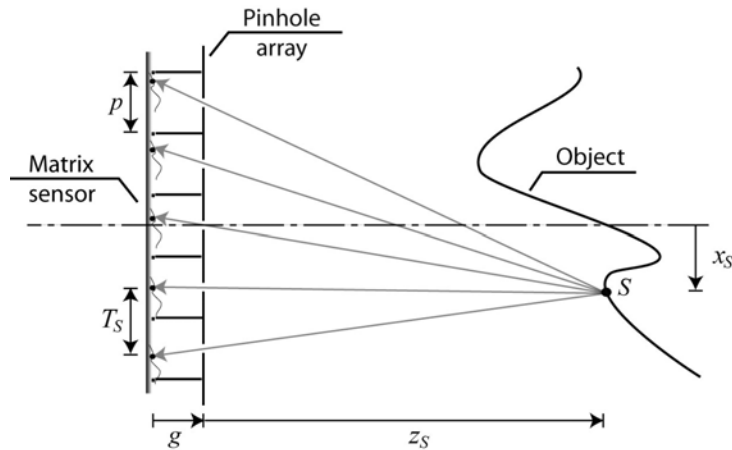


Figure 2. Procedure for recording the integral image through a pinhole array. Each pinhole produces an elemental image with different perspective information.

Let consider, next, an arbitrary point, S , on the surface of a 3D scene placed in front of the pinhole array. The position of such point is

$$\mathbf{x}_S = (x_S, z_S) . \quad (2.)$$

The light rays emitted by such point pass through the pinholes and produce an array of microimages onto the film. The position of the m^{th} microimage is given by

$$\mathbf{x}'_{S,m} = (M_S x_S + m T_S, -g) , \quad (3.)$$

where the parameters M_S and T_S are defined, respectively, as the pickup magnification and period. The pickup magnification can be calculated as

$$M_S = -\frac{g}{z_S} . \quad (4.)$$

The pickup period is the distance between microimages, namely

$$T_S = \frac{z_S + g}{z_S} p . \quad (5.)$$

The number of replica recorded by the system is limited by the opaque barriers, so that only parts of microimages verifying

$$|x'_{S,m} - m p| < \frac{p}{2} . \quad (6.)$$

are registered. This condition limits the maximum (m^+) and minimum (m^-) values of index m for each source point, so that the overall number, n_S , of elemental images produced by S is

$$n_S = m_S^+ + m_S^- + 1 . \quad (7.)$$

The optical barriers limit, as well, the visual field recorded in any elemental image. As shown in Figure 2, the elemental visual field, Φ_p , depends on the object plane depth through

$$\Phi_p(z) = \frac{z}{g} p , \quad (8.)$$

The total field recorded in the film is the sum of an elemental field and the width of the pinhole array, namely

$$\Phi_T(z) = (2M - 1)p + \Phi_p(z). \quad (9.)$$

Consider now that the properly processed photographic film is placed in front of a second pinhole array, the reconstruction array, with pitch p_R . The distance between the film and the array is g_R , and the film is illuminated so that it behaves as a spatially incoherent source.

In Figure 3 we have represented the reconstruction stage. When the geometrical parameters are the same as the ones used in the pickup, symmetrical projection, the emerging rays intersect at a point R that is just at the same position as the object S . However, when the geometrical parameters, g_R , T_R and p_R , are different, the emerging rays no longer intersect at the object position. However, they still intersect at a different position R of coordinates

$$x_R = \frac{p_R}{T_R - p_R} x'_{R,1}, \quad (10.)$$

$$z_R = \frac{p_R}{T_R - p_R} g_R. \quad (11.)$$

The question that comes out at this context is: is it possible the reconstruction of the scene with a given magnification, but without any distortion? To answer, it is useful make the following definition of geometrical magnifications

$$M_p = \frac{p_R}{p}, \quad M_g = \frac{g_R}{g} \quad y \quad M_T = \frac{T_R}{T_S}. \quad (12.)$$

Now, Eq. (10) can be rewritten as

$$x_R = \frac{M_p p}{M_T T_S - M_p p} M_T x'_{S,1} = M_T \frac{T_S - p}{\frac{M_T}{M_p} T_S - p} x_S. \quad (13.)$$

From this equation we can define the following lateral magnification

$$M_x \equiv \frac{x_R}{x_S} = \frac{M_T M_p}{M_T g + (M_T - M_p) z_S} g, \quad (14.)$$

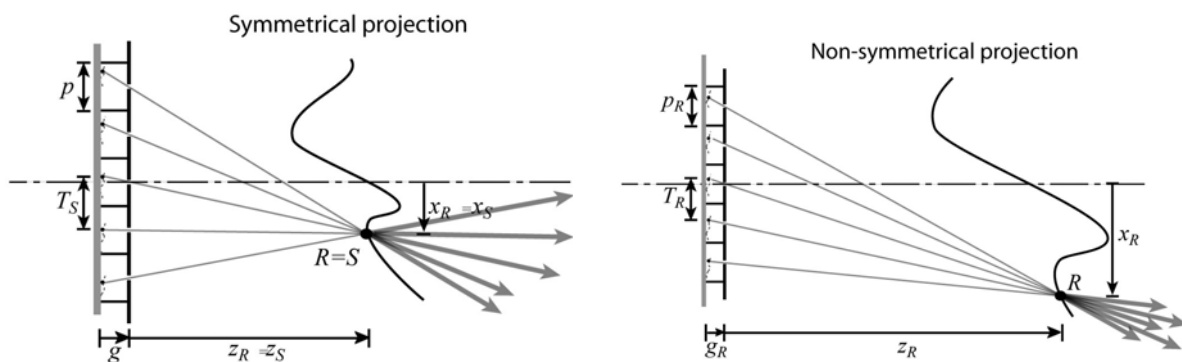


Figure 3. The non-symmetrical projection provides a distorted reconstruction of the 3D scene.

which relates, for each z_S , the lateral dimension of the reconstructed scene with that of the original one. It is apparent that only in case of $M_T=M_p$ the lateral magnification does not depend on z_S .

By use of Ecs. (5) and (12), Eq. (11) can be rewritten as

$$z_R = \frac{M_p p M_g g}{M_T T_S - M_p p} \frac{T_S - p}{p g} z_S = M_g \frac{T_S - p}{\frac{M_T}{M_p} T_S - p} z_S, \quad (15.)$$

so that one can define the axial magnification of the reconstruction as

$$M_z \equiv \frac{z_R}{z_S} = \frac{M_g M_p}{M_T g + (M_T - M_p) z_S} g. \quad (16.)$$

Again, we find that it is necessary that $M_T=M_p$ to obtain a magnification independent of z_S . In such case, $M_z=M_g$. Consequently, the necessary condition for obtaining a magnified but non-distorted reconstruction is $M_g=M_p=M_T$.

2.2. Microlens array

In the Figure 4 we have illustrated the pickup process when performed with a microlens array (MLA). When comparing this system with the one represented in Figure 2 we find that, since the rays passing through the centre of a microlens do not deviate, the integral image is similar to the one obtained with the pinholes. Note however that, due to the imaging property of the lenses, the rays drawn in previous section have been replaced by ray bundles. Consequently, only one plane of the object scene, the object reference plane (ORP), is imaged sharply. The depth, d , of the ORP is calculated through

$$\frac{1}{d} + \frac{1}{g} = \frac{1}{f}. \quad (17.)$$

In Figure 5 we show a scheme of symmetrical projection, which permits the reconstruction of the object scene by intersection of ray cones. Note that although all the cones focus at the image reference plane (IRP), the image is reconstructed at the object positions by intersection of the ray cones. Due to the cones spread, the lateral resolution obtained with the microlenses is worse than the one obtained

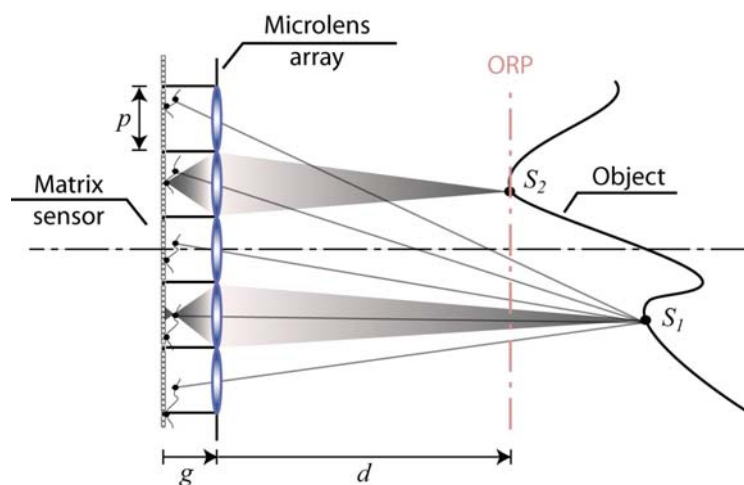


Figure 4. Formation of the integral image through an array of microlenses.

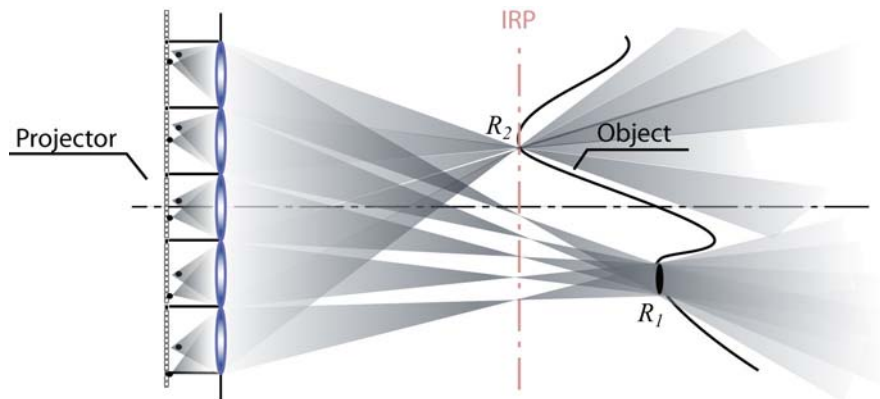


Figure 5. The image is reconstructed by ray cones intersection at the object positions.

with the pinholes.

3. Multifaceted structure of integrated images

The focusing properties of the microlenses strongly influence the structure of the observed reconstructed image. To understand this influence one should take into account that each microlens forms onto the IRP an intermediate image of its corresponding elemental image [16]. The depth of the IRP, d_R , is determined by the projection gap, g_R , and the microlenses focal length, f_R . Such intermediate images are seen by the observer, that is, are imaged onto the observer retina by the eye optics. The image that is formed in the retina (or in the CCD if the observer is a digital camera) will be named here as the integrated image. Due to the limited size of the eye entrance pupil, the observer only see one region, or facet, of each elemental intermediate image, so that the integrated image is formed by the arrangement of the facets observed through each microlens. This fact gives rise to an integrated image with faceted structure.

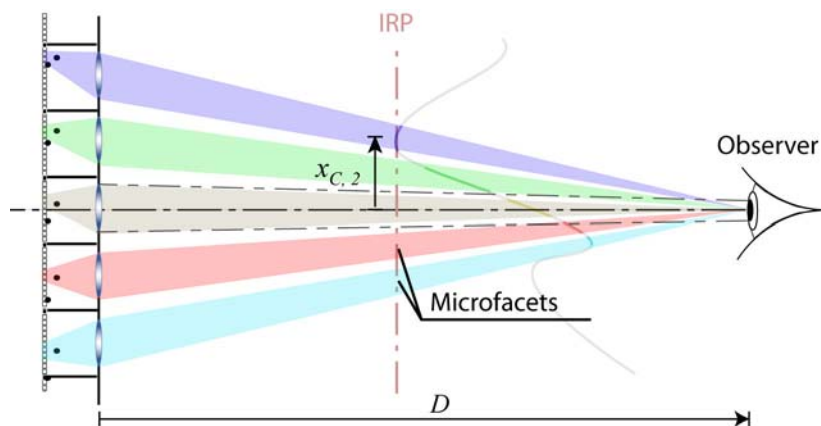


Figure 6. The integrated image is composed by an arrangement of elemental facets, whose shape is determined by the size and position of the eye pupil.

To analyze this phenomenon, let us suppose that the eye is placed at a distance D from the MLA. As seen in the Figure 7, the facets are arranged in rectangular grid and centered at

$$x_{c,m} = m \frac{D - d_R}{D} p_R . \quad (18.)$$

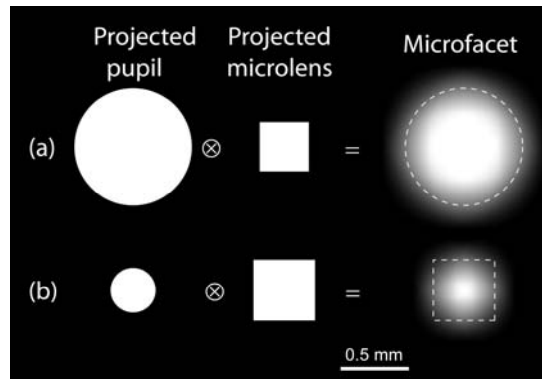


Figure 7. Illustration of calculation of the facet function. (a) When $D=350$ mm microlenses, the facets have a circle-like shape; (b) when $D=900$ mm the facets are square like. In both cases we have marked the form the field of half-illumination (dashed line).

The facets shape can be calculated as the convolution between two projected pupils: (a) the projection, through the eye-pupil center, of the lenslet pupil onto the IRP and (b) the projection, through the lenslet center, of the eye pupil onto the IRP. Then the elemental facets are expressed as

$$E(x, y) = \text{rect}\left(\frac{x}{w}, \frac{y}{w}\right) \otimes \text{circ}\left(\frac{\sqrt{x^2 + y^2}}{R_{obs}}\right), \quad (19.)$$

where

$$R_{obs} = \frac{d_R}{2D} \phi_{obs} \quad \text{and} \quad w = \frac{D - d_R}{D} L, \quad (20.)$$

ϕ_{obs} being the observer pupil diameter, and L the lenslet width. Since the facets are obtained as the result of the convolution between two binary functions, they are no longer binary. Specifically, they consist of a central zone with uniform illumination and an outer zone of vignetting where the illumination falls off gradually. For distances D larger than the distance D_o , defined as

$$D_o = d_R \frac{\phi_{obs} + L}{L}, \quad (21.)$$

the eye pupil acts as the exit pupil of the system, and the lenslet acts as the exit window. Thus the facets have a square-like shape. In contrast, for eye positions such that $D < D_o$, the eye pupil acts as the exit window, and therefore the facets have a circle-like shape. In Figure 6 we show the facets corresponding to two different values of D . For calculations we assumed a typical integral-image setup, that is $\phi_{obs}=3$ mm, $d_R = 100$ mm, $p_R = 1$ mm, and $f_R = 3.3$ mm.

To illustrate the multifacet phenomenon, we gave performed an hybrid experiment in which the integral image is obtained experimentally in the laboratory, whereas the integrate image has been calculated by computer. In the experiment the 3D scene consisted in two letters, G and R, printed in transparent plates and placed at 100 mm and 50 mm from the MLA, respectively.

A MLA composed by 39 x 27 square microlenses of 1.01 mm width and $f=3.3$ mm. The images were captured with a Canon EOS 350D camera equipped with macro objective Canon 60 mm. The CMOS sensor was composed by 3455 x 2304 pixels. In Figure 8 we show the integrated images calculated for four different configurations. In (a), (b) and (d) the quality of the integrated image is very low due to un-efficient arrangement of the facets. Only in the image (c) the two letters appear in focus and without any pattern. In view if these images one can conclude that the use of MLA with fill

factor equal to one is necessary in the reconstruction. Besides, observation distances longer than D_0 are convenient to avoid the patterned image.

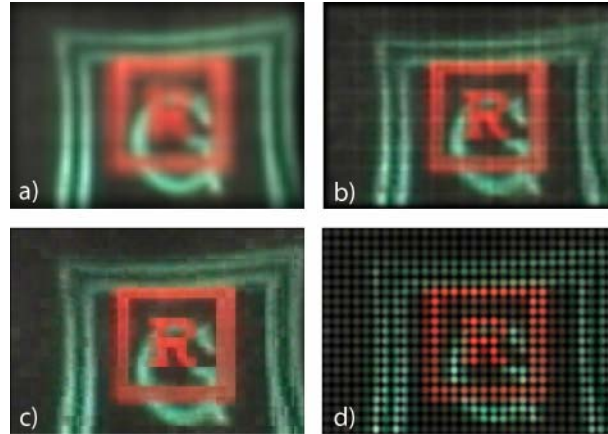


Figure 8. Integrated images obtained assuming that for observer: (a) $\phi_{\text{obs}}=20$ mm and $D=400$ mm; (b) $\phi_{\text{obs}}=10$ mm and $D=400$ mm; (c) $\phi_{\text{obs}}=3$ mm and $D=1000$ mm; (d) same but with filling factor equal to 0.5.

Acknowledgment

This work has been funded in part by the Plan Nacional I+D+I (grant DPI2006-8309), Ministerio de Ciencia y Tecnología, Spain. R. Martínez-Cuenca acknowledges funding from the Universitat de València (Cinc Segles grant).

References

- [1] Smith R 1738 *A Complete System of Optics in Four Books, viz. A popular, a mathematical, a mechanical, and a philosophical treatise. To which are added remarks upon the whole.* (Cambridge) **1** 280, **2** 281–455
- [2] Porterfield W 1759 *A Treatise on the Eye, the Manner and Phaenomena of Vision* (Edinburgh: Hamilton and Balfour) 2 volumes
- [3] Wheatstone C 1838 *Philosophical Transactions on Royal Society of London* **128** 371–94
- [4] Rollmann W 1945 *Pogg. Ann.* **1853** 90
- [5] Lippmann G 1908 *Journal of Physics (Paris)* **7** 821–825
- [6] Ives H E 1928 *J.O.S.A. & R.S.I.* **17** 435–39
- [7] Ives H E 1931 *J. Opt. Soc. Am.* **21** 171–76
- [8] Ives H E 1931 *J. Opt. Soc. Am.* **21** 397–409
- [9] Burckhardt C B 1968 *J. Opt. Soc. Am.* **58** 71–6
- [10] Okoshi T, Yano A and Fukumori Y 1971 *Appl. Opt.* **10**, 482–89
- [11] Okoshi T 1971 *Appl. Opt.* **10** 2284–91
- [12] Yang L, McCornick M and Davies N 1988 *Appl. Opt.* **27** 4529–34
- [13] Okano F, Hoshino H, Arai J and Yuyama I 1997 *Appl. Opt.* **36** 1598–603
- [14] Arai J, Okui M, Kobayashi M and Okano F 2004 *J. Opt. Soc. Am. A* **21** 951–8
- [15] Arai J, Okano F, Hoshino H and Yuyama I. 1998 *Appl. Opt.* **37** 2034–45
- [16] Martínez-Corral M, Javidi B, Martínez-Cuenca R and Saavedra G 2005 *J. Opt. Soc. Am. A* **22** 597–603