Extended depth-of-field in Integral Imaging by depth-dependent deconvolution

H. Navarro^{*1}, G. Saavedra¹, M. Martinez-Corral¹, M. Sjöström², R. Olsson², ¹Dept. of Optics, Univ. of Valencia, E-46100, Burjassot, Spain. ²Dept. of Information Technology and Media, Mid Sweden Univ., Sundsvall, Sweden.

ABSTRACT

Integral Imaging is a technique to obtain true color 3D images that can provide full and continuous motion parallax for several viewers. The depth of field of these systems is mainly limited by the numerical aperture of each lenslet of the microlens array. A digital method has been developed to increase the depth of field of Integral Imaging systems in the reconstruction stage. By means of the disparity map of each elemental image, it is possible to classify the objects of the scene according to their distance from the microlenses and apply a selective deconvolution for each depth of the scene. Topographical reconstructions with enhanced depth of field of a 3D scene are presented to support our proposal.

Keywords: Integral Imaging, depth of field, depth map, deconvolution.

1. INTRODUCTION

At present there is a wide variety of methods to obtain 3D images. Depending on whether or not the observer needs to wear special glasses to perceive the 3D sensation we can differentiate between stereoscopic or autostereoscopic systems. Among the techniques that belong to the latter group, we find Integral Imaging systems. These systems work with incoherent illumination and hence can capture and display true color 3D images [1]. Integral Imaging provides stereo parallax as well as full and continuous motion parallax, allowing multiple viewing positions for several viewers. It is a relatively old concept which was initially proposed by Gabriel Lippmann in 1908 under the name Integral Photography [2]. The Lippmann idea was placing a flat sensor behind an array of microlenses so that each lenslet images a different perspective of a 3D scene over the sensor. The 3D scene can be reconstructed by projecting the recorded 2D elemental images on a flat display placed in front of another microlens array of the same characteristics. The range of possible viewing angles, both horizontally and vertically, is given by the numerical aperture of each lenslet. Larger numerical apertures provide larger viewing angles, but as the numerical aperture increases, the depth of field (DOF) of the system becomes smaller. A limited DOF is a serious problem for this kind of systems because to reconstruct clear 3D images it is essential to capture sharp 2D elemental images. Objects belonging to the 3D scene located in axial positions out of the DOF will appear blurred in the elemental image, and hence, will also be blurred in the reconstructed image. Some methods have been proposed to increase the DOF of an Integral Imaging system [3]. In this paper we propose a new method to extend the DOF of Integral Imaging systems in the reconstruction stage. The idea is based in reversing the out-of-focus blur by a selective depth-dependent deconvolution. First we get the depth map for each elemental image to obtain the depth information of the 3D scene. The axial range spanning the scene is divided into intervals and then each elemental image is filtered by selecting the pixels that are associated with a given interval. It is possible to define an effective point spread function (PSF) in good approximation over each interval of distances. Only the pixels belonging to a certain interval are deconvolved with the effective PSF calculated for the center of the interval. The final image is the sum of the pixels of every interval after being filtered and deconvolved. We have simulated topographical reconstructions from the elemental images captured from a real scenario to show the ability of this technique to extend the DOF of an Integral Imaging system in the reconstruction stage.

*hector.navarro@uv.es; phone +34963544042; www.uv.es/imaging3

Stereoscopic Displays and Applications XXIV, edited by Andrew J. Woods, Nicolas S. Holliman, Gregg E. Favalora, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8648, 86481H · © 2013 SPIE-IS&T CCC code: 0277-786X/13/\$18 · doi: 10.1117/12.2013705

2. DIFFRACTIVE ANALYSIS OF THE CAPTURE STAGE

We will start by describing the capture stage from the point of view of diffraction theory. Instead of using a microlens array we have use the so-called Synthetic Aperture Method [6] in which all the elemental images are picked up with only one digital camera that is mechanically translated. Usually, camera lens is an assembly of lenses and a multi-leaf diaphragm. These elements are embedded inside a cover and some of them can move to allow focusing at different depths. The use of many lens elements is designed to minimize aberrations and to provide sharp images. Therefore, calculate the PSF of a camera lens would be a complex task if we had to take into account all its component lenses and their exact positions. As an alternative, we propose a much simpler method in which we only need to know some parameters that can be easily measured. Once the camera is focusing on a particular plane, to obtain the theoretical PSF we just need to know the diameter of the entrance pupil of the camera lens, its distance to the in-focus plane and the magnification factor between the in-focus plane and the sensor plane. To have an intuitive understanding of our method is convenient to analyze first the following configuration in which we have outlined a scheme for a generic camera lens with an arbitrary f-number.



Figure 1. The figure shows the lens elements, the position of the variable leaf diaphragm, and the image of its opening formed by the elements in front of it. This image is the entrance pupil of the camera lens.

A useful concept to determine which rays will transverse the entire optical system is the entrance pupil. It can be considered as the window through which light enters the objective. The entrance pupil of the camera lens is the image of the diaphragm as seen from an axial point on the object through those elements preceding the stop. Suppose that the camera shown in Fig. 1 is focusing on a reference plane located at a distance α from the entrance pupil of the camera lens. Because the diaphragm is a circular stop, the entrance pupil has also circular shape.



Figure 2. Scheme of the capture setup. All rays passing through the entrance pupil traverse the entire optical system and reach the sensor. Object points out of the reference plane produce blurred images in the sensor.

Now we will consider the light scattered at an arbitrary point (x_0, y_0, z_0) of the surface of a 3D object (see Fig. 2). For simplicity we assume quasi-monochromatic illumination with mean wavelength λ . Spatial coordinates are denoted (x, y) and z for directions transverse and parallel to the system main optical axis. The amplitude distribution of the spherical wave emitted by that point in the plane of the entrance pupil can be written as

$$U_E^{\lambda}(x,y;z) = -\frac{1}{z+\alpha} e^{-ik(z+\alpha)} \exp\left\{-i\frac{k}{2(z+\alpha)}(x^2+y^2)\right\} p_E(x,y)$$
(1)

where k is the wave number $k = 2\pi/\lambda$. In this equation we have multiplied the impinging wave-front by the amplitude transmittance of the entrance pupil, $p_E(x, y)$. In order to simplify resulting equations, we will use a mathematical artifice consisting on a virtual propagation from the plane of the entrance pupil to the reference plane. The virtual amplitude distribution in the reference plane is given by

$$U_{R}^{\lambda}(x,y;z) = \frac{e^{ikz}}{(z+\alpha)\alpha} \exp\left\{-i\frac{k}{2\alpha}(x^{2}+y^{2})\right\} \int_{-\infty}^{+\infty} p_{E}(x_{0},y_{0}) \exp\left\{-i\frac{k}{2}\frac{z}{(x+\alpha)\alpha}(x_{0}^{2}+y_{0}^{2})\right\} \\ \exp\left\{i\frac{2\pi}{\lambda\alpha}(xx_{0}+yy_{0})\right\} dx_{0} dy_{0}$$
(2)

Since the reference plane and the camera sensor are conjugated through the camera lens, the amplitude distribution over the sensor is a scaled version of the amplitude distribution at the reference plane. The scaling factor is given by the lateral magnification, M_0 , between these two planes.

$$U_{S}^{\lambda}(x, y; z) = \frac{1}{M_{0}} U_{R}^{\lambda} \left(\frac{x}{M_{0}}, \frac{y}{M_{0}}; z \right)$$
(3)

Introducing Eq. (2) into Eq. (3), we straightforwardly find that

$$U_{S}^{\lambda}(x,y;z) = \frac{e^{ikz}}{M_{0}(z+\alpha)\alpha} \exp\left\{-i\frac{k}{2M_{0}^{2}\alpha}(x^{2}+y^{2})\right\} \int_{-\infty}^{+\infty} p_{E}(x_{0},y_{0}) \exp\left\{-i\frac{k}{2}\frac{z}{(x+\alpha)\alpha}(x_{0}^{2}+y_{0}^{2})\right\} \\ \exp\left\{i\frac{2\pi}{\lambda M_{0}\alpha}(xx_{0}+yy_{0})\right\} dx_{0} dy_{0}$$
(4)

If we use a local reference system that moves with the camera, impulse response has radial symmetry around the optical axis of the camera lens. Therefore using cylindrical coordinates is best suited to write our equations. Accordingly, Eq. (4) can be rewritten as

$$U_{\mathcal{S}}^{\lambda}(r;z) = \frac{2\pi e^{ikz}}{M_0(z+\alpha)\alpha} \exp\left\{-i\frac{k}{2M_0^2\alpha}r^2\right\} \int_0^{\emptyset_E/2} p_E(\rho) \exp\left\{-i\frac{\pi}{\lambda}\frac{z}{(z+\alpha)\alpha}\rho^2\right\} J_0\left(\frac{2\pi}{\lambda\alpha M_0}r\rho\right)\rho d\rho$$
(5)

where ϕ_E is the diameter of the entrance pupil of the camera lens and *r* is the radial coordinate over the sensor plane. The response generated in the sensor of the camera is proportional to the intensity of the incident light. Therefore, the intensity impulse response can be obtained as the squared modulus of the function in Eq. (5)

$$H_{S}^{\lambda}(r;z) = \left| U_{S}^{\lambda}(r;z) \right|^{2} \tag{6}$$

Function $H_S^{\lambda}(r; z)$ has a strong dependence on the axial position of the corresponding surface points. Consequently, the impulse response is different at any depth. Therefore, the PSF cannot be rigorously defined. However, the impulse response can be regarded as the sum of the impulse responses generated by a continuum of point sources axially distributed. This fact can be written as

$$H_{S}^{\lambda}(r,z) = \int H_{S}^{\lambda}(r,\epsilon)\delta(z-\epsilon)d\epsilon$$
⁽⁷⁾

From this interpretation, it is possible define the intensity PSF for each depth with respect to the reference plane, which is precisely $H_S^{\lambda}(r, \epsilon)$. Given a plane at a distance $z = \epsilon$ from the reference plane, $\Pi_{\epsilon}(r, \theta)$, its image over the sensor can be expressed as the 2D convolution of a scaled version of the intensity distribution scattered at that plane, $\Pi_{\epsilon}(r/M_{\epsilon}, \theta)$, and the function $H_S^{\lambda}(r, \epsilon)$.

$$I_{S}^{\lambda,\epsilon} = \frac{1}{M_{\epsilon}^{2}} \Pi_{\epsilon} (\frac{r}{M_{\epsilon}}, \theta) \bigotimes_{r,\theta} H_{S}^{\lambda}(r,\epsilon)$$
(8)

The scaling factor M_{ϵ} comes from the lateral magnification of the camera lens and depends on the distance ϵ from the plane of interest to the reference plane.

SPIE-IS&T/ Vol. 8648 86481H-3

We will extend now our analysis to a volume object. Suppose a photo camera whose sensor is conjugated through the camera lens with some plane cutting the object in two parts. The intensity distribution of incoherent light scattered by the object surface can be represented by the real and positive function $O(r, \theta, z)$. This function can be expressed as

$$O(r,\theta,z) = \int O(r,\theta,\epsilon) \,\,\delta(z-\epsilon)d\epsilon \tag{9}$$

which can be interpreted as if we had sliced the object in infinitesimal width sheets. From Eq. (8), the image of one of these sheets over the sensor can be written as

$$I_{S}^{\lambda,0_{\epsilon}}(r,\theta) = \frac{1}{M_{\epsilon}^{2}}\delta(z-\epsilon)O\left(\frac{r}{M_{\epsilon}},\theta,z\right)\bigotimes_{r,\theta}H_{S}^{\lambda}(r,\epsilon)$$
(10)

The image of the volume object over the sensor can be considered as the sum of the images of each slice of the object over the sensor. The intensity distribution in that plane is given by

$$I_{S}^{\lambda,0}(r,\theta) = \frac{1}{M_{\epsilon}^{2}} \int \delta(z-\epsilon) O\left(\frac{r}{M_{\epsilon}},\theta,z\right) \bigotimes_{r,\theta} H_{S}^{\lambda}(\rho,\epsilon) d\epsilon$$
(11)

After a careful analysis of Eq. (11) we see that once the camera is focused at a given distance, at the sensor plane appear, together with the in-focus image of the reference plane, the blurred images of the rest of sections that constitute the 3D object. Let us take a slice of the object in a plane perpendicular to the optical axis. As stated in Eq. (10), the image of this slice over the sensor can be expressed as the 2D convolution of the intensity distribution at that plane with the intensity PSF associated with that depth. By knowing the PSF for that depth, it is possible to reverse out-of-focus blur which was made by the optical system. There are many methods to do this, but in order to minimize the impact of the photon noise in the restored image we will use Richardson-Lucy deconvolution [7]. The algorithm is an iterative procedure based on maximizing the likelihood of the resulting image. Recovered image is considered an instance of the original image under Poisson statistics.

3. DEPTH EXTRACTION AND DECONVOLUTION

Deconvolution tools cannot be applied to recover simultaneously a sharp version of objects located at different depths if that objects are affected from a different blur. However, if we are able to identify on each elemental image which depth in the scene is associated with each pixel of the image, we could select the pixels associated with the surfaces located at a given depth. This being so, the axial range that spans the 3D scene is divided into intervals, and the pixels of each elementary image are classified according to the axial interval with which they are associated. In order to do this, we propose obtaining the depth map of each elemental image. An integral image may be considered as a set of stereo pairs so that we can use well developed stereo vision algorithms to get the depth information. The output of the stereo computation is a disparity map which tells how far each point in the physical scene was from the camera. The heart of any stereo vision system is stereo matching, the goal of which is to establish correspondence between two points arising from the same element in the scene. Stereo matching is usually complicated by several factors such as lack of texture, occlusion, discontinuity and noise. Basic block matching chooses the optimal disparity for each pixel based on its own cost function alone, but this technique creates a noisy disparity image [9]. To improve these results we have used the optimization method proposed by Veksler [10] in which the disparity estimation at one pixel depends on the disparity estimation at the adjacent pixels. The maximum number of distances that we can distinguish is given by the number of bits used in the disparity map. For example, if we are working with 8 bit depth map, we can distinguish 256 distances in the range covering the scene. But for our purpose, it is usually not necessary to use all these intervals. As the complexity of the surfaces of the objects composing the 3D scene increases, we must increase the number of axial intervals in which we divide the scene. Overlooking the deconvolution process, it is useful to approximate the intensity distribution of the 3D objects using constant depth segment as

$$O(x, y, z) = O(x, y, z_i), \text{ for } z_i \le z < z_{i+1}, \quad i = 1, \dots, N$$
(12)

where the object is confined between z_0 and z_{N+1} . Only the pixels belonging to a certain interval are deconvolved in the elemental image with the effective PSF calculated for that interval.

4. IMPROVEMENT OF THE DOF IN THE RECONSTRUCTION STAGE

To validate the proposed method we performed a hybrid experiment in which the elemental images were captured experimentally and the reconstruction was simulated computationally. In Fig. 3 we show the experimental setup for the pickup stage. As we can see in the picture, the scene was composed by a wooden background and three resolution charts located at different distances from the camera.



Figure 3. Experimental setup for the acquisition of the elemental images of a 3D scene.

A digital camera is mechanically translated in a rectangular grid in steps of 5 mm both in the horizontal and vertical direction. It was accurately calibrated so that the optical axis was perpendicular to the plane defined by the grid in which the camera is moved. A set of 9x9 elemental images of 2000x1864 pixels each was taken with a camera lens of focal length f=29 mm. The sensor of the camera was conjugated with the wooden surface in the background which was located at a distance $\alpha=710.1 \text{ mm}$ of the entrance pupil of the photographic lens. The f/number of the camera lens was chosen to f#=4.5, and the lateral magnification for the in-focus plane was $M_0=0.053$. Finally, the entrance pupil of the camera lens was measured with the help of a microscope, obtaining a diameter $\phi_E=9.8 \text{ mm}$. All geometric parameters of the setup are known so it is straightforward obtaining the distance of the objects composing the image to the in-focus plane from the disparity information provided by the depth map. The method mentioned in section 3 was applied to the various stereo pairs composing the integral image in order to get the depth map for each elemental image. As an example, the depth map obtained for one of the captured elemental images is presented below



Figure 4. Elemental image with shallow DOF (Left). Depth map for that elemental image (Right).

SPIE-IS&T/ Vol. 8648 86481H-5

Given the distance α of the entrance pupil to the in-focus plane, it is straightforward obtaining the distance z of each surface of the scene to the reference plane from the depth map information. As stated in the previous section, depending on the complexity, the scene is divided in intervals. Our scene is mainly composed by three flat objects located at different depths plus a background and therefore the scene can be divided in four intervals. According to Eq. (6), the intensity PSF associated with each interval can be calculated theoretically. For each elemental image, pixels associated with a given interval are deconvolved with its corresponding intensity PSF. In Fig. 5 we can see the result of this process for the elemental image in Fig. 4.



Figure 5. Filtering and deconvolution of each axial interval.

The final elemental image with extended DOF is obtained as the sum of the pixels of every interval after being filtered and deconvolved. Figure below shows the elemental image of Fig. 4 with extended DOF.



Figure 6. Elemental image with extended DOF.

In Fig. 6 we can appreciate two phenomena that lead to a poor visual aspect in the extended DOF elemental image. The ringing occurs along the areas of sharp intensity contrast in the image during the deconvolution process and abrupt transitions are due to the border of the areas selected in the depth map for each axial interval. These phenomena are not a problem when we are interested in performing topographical reconstructions of the 3D scene. The back-projection technique described in [11] is used to reconstruct the scene focused at different depths. Each elemental image is back-projected on the desired reconstruction plane through a virtual pinhole array. The collection of all back-projected elemental images is then superimposed computationally to achieve the intensity distribution on the reconstruction plane [12]. Ringing effects and abrupt transitions are averaged and smoothed, improving the visual appearance of the images reconstructed at different depths.

In the next figure we show a collection of reconstructions of the scene for the depths where the resolution charts where located. A comparison of the reconstructions obtained with and without applying our method is presented.



Figure 7. Computational reconstructions at different depths from the captured elemental images (a, b, c). Computational reconstructions at different depths after applying the proposed method to each elemental image (d, e, f).



Figure 8. Enlarged version of the areas enclosed by the red line in Fig. 7.

In Fig. 7 we can see the results of the reconstruction in the planes of the resolution charts. Parts of the scene that are out of the DOF of the camera appear blurred on each elemental image. As the distance of the objects to the in-focus plane increases they suffer from increasing blur. This blur produces deterioration in the lateral resolution of the reconstructed images because objects that were captured blurred appear also blurred in the reconstruction. From left to right of the upper row of Fig. 8, we can see how the blur increases as the distance of the resolution chart to the in focus plane becomes larger. In the lower row of the same figure, we show the reconstruction of the same resolution charts at the same depths after applying our method to each captured elemental image. Comparing both rows, it is easy to see that in the resolution charts of the lower row we can resolve frequencies that in the upper row are impossible to distinguish. These results prove the ability of our method to extend the DOF in the reconstructions stage.

5. CONCLUSIONS

We present a method to extend the DOF of an Integral Imaging system in the reconstruction stage. A set of elemental images with shallow DOF was captured experimentally. The technique is based in the use of the depth information provided by the disparity map of each elemental image. Only pixels of the elemental image associated to a certain depth interval are deconvolved with the effective PSF calculated for that segment. The final elemental image with extended DOF is the sum of the pixels related with each interval after being filtered and deconvolved. Using a back-projection algorithm, we have simulated topographical reconstructions of the 3D scene from the captured elemental image. These reconstructions have been compared with those obtained after applying out method to each elemental image. We have recovered frequencies of the objects reconstructed at different depths which without applying the proposed technique cannot be resolved.

6. ACKNOWLEDGMENTS

This work was supported in part by the Plan Nacional I+D+I under Grant DPI2012-32993, Ministerio de Economía y Competitividad. Héctor Navarro gratefully acknowledges funding from the Generalitat Valencia (VALi+d predoctoral contract).

7. REFERENCES

- [1] B. Javidi and F. Okano, eds., [Three-dimensional television, video and display technologies], Springer-Verlag, (2002).
- [2] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," J. Phys. 7, 821-825 (1908).
- [3] R. Martínez-Cuenca, G. Saavedra, M. Martínez-Corral and B. Javidi, "Enhanced depth of field integral imaging with sensor resolution constraints," Opt. Express 12, 5237-5242 (2004).
- [4] M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca and G. Saavedra, "Integral imaging with improved depth of field by use of amplitude-modulated microlens array," Appl. Opt. 43, 5806-5813 (2004).
- [5] A. Castro, Y. Frauel, and B. Javidi, "Integral imaging with large depth of field using an asymmetric phase mask," Opt. Express 15, 10266-10273 (2007).
- [6] J.-S. Jang and B. Javidi, "Three-dimensional synthetic aperture integral imaging," Opt. Lett. 27, 1144-1146 (2002).
- [7] William H. Richardson, "Bayesian-Based Iterative Method of Image Restoration," J. Opt. Soc. Am. 62, 55-59 (1972).
- [8] L. B. Lucy, "An iterative technique for the rectification of observed images," AJ 79, 745-754 (1974).
- [9] E. Trucco and A. Verri, [Introductory Techniques for 3-D Computer Vision], Prentice Hall, (1998).
- [10] O. Veksler, "Stereo Correspondence by Dynamic Programming on a Tree," Proc. IEEE 2, 384-390 (2005).
- [11] S.-H. Hong, J.-S. Jang, and B. Javidi, "Three-dimensional volumetric object reconstruction using computational integral imaging," Opt. Express 12, 483-491 (2004).
- [12] H. Navarro, G. Saavedra, A. Molina, M. Martinez-Corral, R. Martinez-Cuenca, and B. Javidi, "Optical slicing of large scenes by synthetic aperture integral imaging," Proc. SPIE 7690, 7690-0M (2010).