Is it worth using an array of cameras to capture the spatio-angular information of a 3D scene or is it enough with just two?

H. Navarro, A. Dorado, G. Saavedra, A. Llavador, and M. Martínez-Corral

Department of Optics, University of Valencia, E-46100 Burjassot, Spain.

Hector.Navarro@uv.es, Adrian.Dorado@uv.es, Genaro.Saavedra@uv.es, Anabel.llavador@uv.es, Manuel.Martinez@uv.es

Bahram Javidi

Electrical and Computer Engineering Dept., University of Connecticut, Storrs, CT 06269-2157

bahram@engr.uconn.edu

ABSTRACT

An analysis and comparison of the lateral and the depth resolution in the reconstruction of 3D scenes from images obtained either with a classical two view stereoscopic camera or with an Integral Imaging (InI) pickup setup is presented. Since the two above systems belong to the general class of multiview imaging systems, the best analytical tool for the calculation of lateral and depth resolution is the ray-space formalism, and the classical tools of Fourier information processing. We demonstrate that InI is the optimum system to sampling the spatio-angular information contained in a 3D scene.

Key words: Integral imaging, 3D imaging and display, 3D resolution

In this contribution we analyze and compare the lateral and the depth resolution in the reconstruction of 3D scenes from images obtained either with a classical two view stereoscopic camera or with an Integral Imaging (InI) pickup setup. Since the two above systems belong to the general class of multiview imaging systems, the best analytical tool for the calculation of lateral and depth resolution is the ray-space formalism, and the classical tools of Fourier information processing.

To explain these tools, let us start by stating that when the points of the surface of a 3D object are, themselves, sources of radiated light or, equivalently, act as secondary sources after been illuminated by a primary one, it is created a flow of light through the space. And this flow of light carries the information about the shape, color and position of the 3D object. This lightfield can be represented through the ray-space function [1]-[2], which is based in the geometrical optics (spatially incoherent illumination, and objects significantly larger than light wavelength). The simplest example is the lightfield generated by point source. Since the source radiate in all directions with the same intensity, the lightfield is represented in the ray space diagram with a straight line (See Fig. 1). As seen in the Figure, we use from the ray-space representation the spatial coordinate and angle of the impact of radiated rays with the reference plane $z=0$

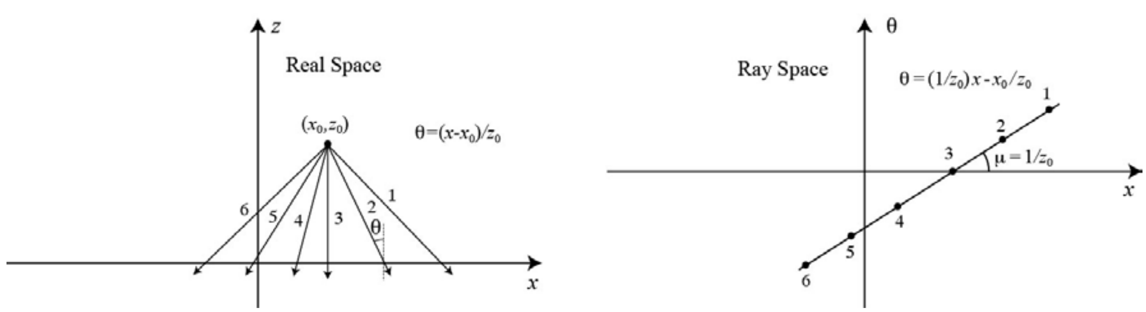

Fig. 1 The lightfield radiated by a point source is represented in the ray-space diagram through a straight line.

Three-Dimensional Imaging, Visualization, and Display 2012, edited by Bahram Javidi, Jung-Young Son, Proceedings of SPIE Vol. 8384, 838406 · © 2012 SPIE
CCC code: 0277-786X/12/$18 · doi: 10.1117/12.923140

Not much more complicate is the ray-space representation of the lightfield radiated by a plane object parallel to the direction of observation. Any point of the object is represented in the ray-space diagram through a straight line of slope μ. The bundle composed by the lines produced by all the points of the plane object constitute the lightfield, as shown in the Figure 2.

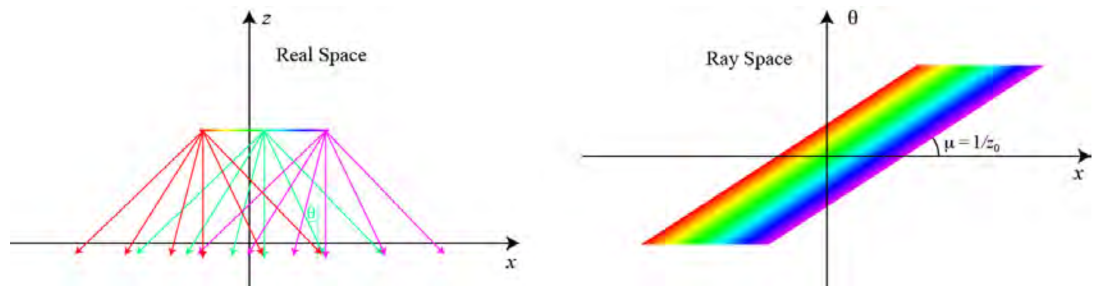

Fig. 2 The lightfield radiated by a plane object is represented in the ray-space diagram through a bundle of slope μ.

In mathematical terms we can express the lightfield generated by a plane object in the following way. Let as assume that we call $f(x)$ the function that describes the light radiance of the object. Then we can built the 2D function

$$g(x,\theta) = f(x)\,\mathrm{rect}\left(\frac{\theta}{\alpha}\right), \tag{1}$$

were α is the maximum ray angle collectible by our system. Then the lightfield, $\ell(x,\theta)$, can be expressed as a sheared version of $g(x,\theta)$, that is

$$\ell(x,\theta) = g(x-\mu\theta,\theta)\ . \tag{2}$$

Naturally, in the general case of a 3D object, the lightfield would be represented in the ray-space through the superposition of bundles with continuously varying slope. Note that the slope of any bundles is proportional to the depth position of the corresponding object section. A very convenient way of analyzing a lightfield is through its spectrum [5], which is obtained by performing the 2D Fourier transform of $\ell(x,\theta)$, that is

$$\Gamma(\Omega_x,\Omega_\theta) = \iint \ell(x,\theta)\exp\left[i2\pi(x\Omega_x+\theta\Omega_\theta)\right]dx\,d\theta\ . \tag{3}$$

As we can see in the Fig 3(a), the spectrum of the lightfield of a flat object is given, basically, by a slit inclined with slope $\mu-\pi/2$. The spectrum of a 3D object is confined to the tie bow obtained by summing up slits with varying slope.

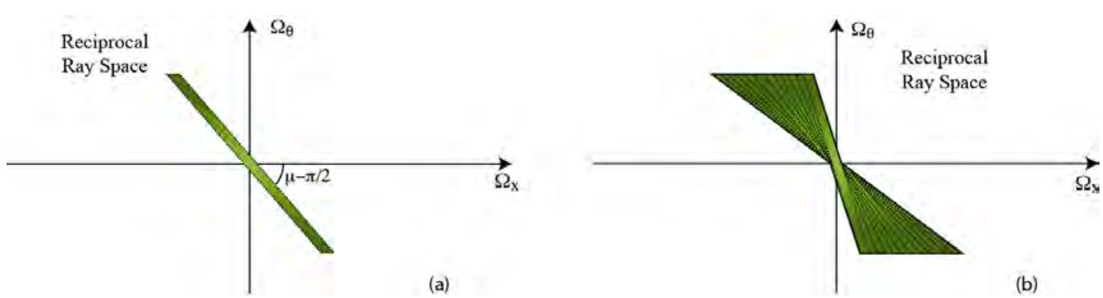

Fig. 3 The spectrum of a lightfield radiated by a 3D object has a bow tie shape.

A very intelligent way of recording information about the lightfield radiated by 3D objects is the use of multi-view camera systems. As shown in Figure 4, an array of equidistant cameras placed at the reference plane ($z=0$) can acquire discrete information about the spatial coordinate and angle of the impact of radiated rays. The image recorded with any of the cameras contains discrete information of the angles of rays passing through one particular spatial coordinate. Thus the pixels of any of the images (from now on: elemental images) contain information corresponding a vertical line in the ray-space diagram. The pixel of the other images, contain information corresponding to parallel, equidistant, vertical lines. In other words, the set of elemental images contain a sampling of the lightfield radiated by the 3D object.

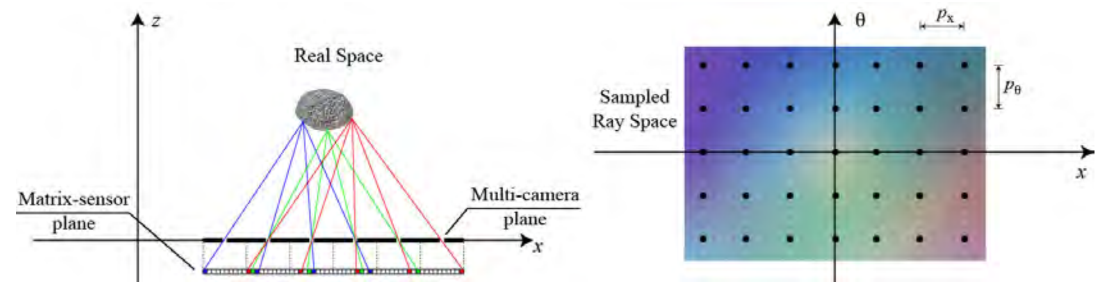

Fig. 4 A multiview camera system captures, and therefore stores, a sampling of the lightfield radiated by 3D objects.

The question that comes out here is the following: Is it possible to recover the continuous lightfield from the sampled one? If so, is it then possible to reconstruct with good axial and depth resolution from the recovered lightfield? To find the answer it is very convenient to reason in terms of the spectrum of the lightfield. According to the sampling theorem [4], the Fourier transform of the lightfield generated by a 3D scene has a form like the one shown in the Figure 5(a). In this Figure we see that the spectrum is composed by a series of equidistant replica of the spectrum of the original lightfield (see Fig. 3). Note that in this case it is possible to apply a rhomboidal window to recover the original lightfield. From this recovered lightfield one can calculate, with good lateral and depth resolution the intensity distribution in the different sections (parallel to the reference plane) of the 3D scene.

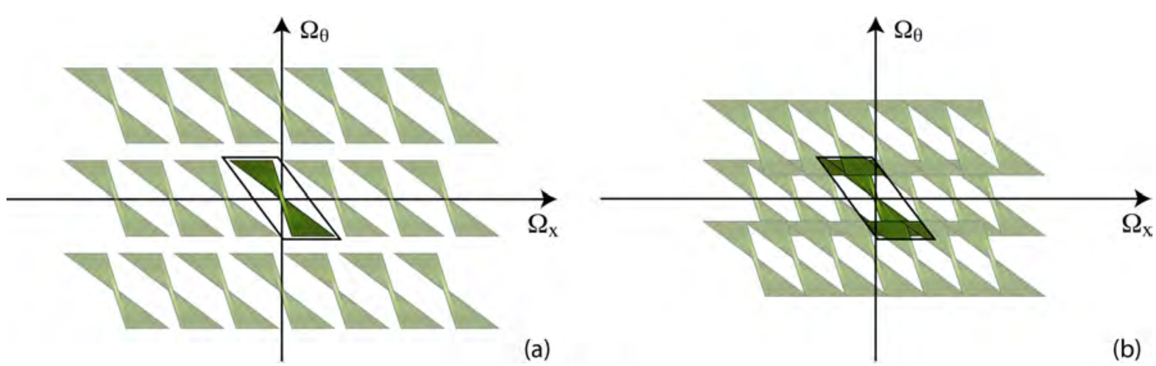

Fig. 5 (a) The spectrum of sampled lightfield when the sampling period is optimum; (b) Same, but when the sampling period is too large.

A great problem appears when the sampling done in the lightfield is not dense enough. In this case the replica of the lightfield spectrum overlap in the reciprocal ray space. Now it is not possible to recover, even in case of using an optimum rhomboidal filter, the original lightfield. This overlapping will be the responsible of the appearance of duplicated ghost images when trying to reconstruct the intensity distribution in the sections of the 3D scene.

Now, after this study of properties of the capture and the calculation of the lightfield, we can find out which kind of multiview system is more adequate for the reconstruction, with good lateral and depth resolution. We compare the typical stereoscopic system (in which a pair of high-resolution pictures of the 3D scene are taken) with the Integral Imaging system (in which many low-resolution pictures are taken). To make the comparison we assume that the total number of pixels is the same in both cases. In the case of an integral imaging system, see Fig. 6, we show that a homogeneous sampling of the lightfield is obtained. This optimum sampling permits to recover the continuous lightfield after a computer filtering process, as shown before in the example of Fig. 5(a).

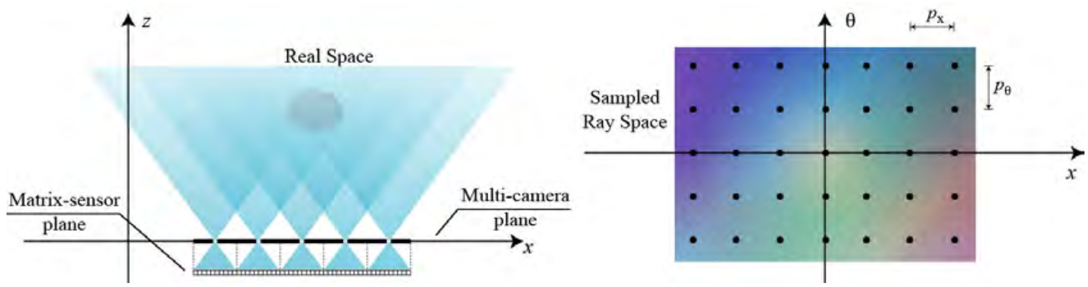

Fig. 6 The integral imaging geometry permits to obtain homogeneous sampling (in both spatial and angular coordinate) of the lightfield.

On the contrary, the stereoscopic camera system performs a very inefficient sampling of the lightfield. The angular information is oversampled, while the spatial information is strongly undersampled (see Fig. 7).

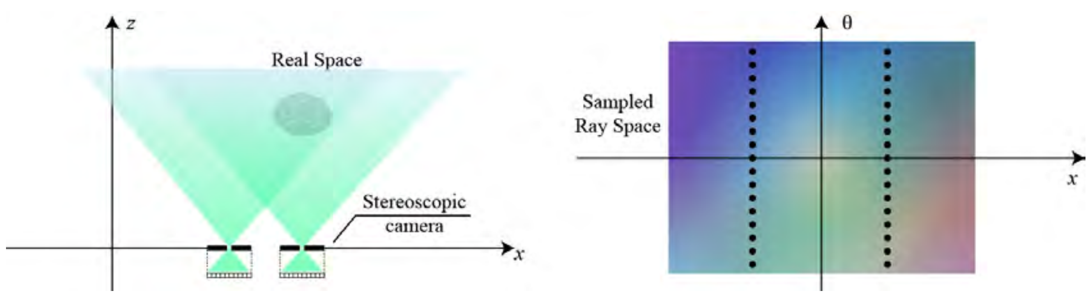

Fig. 7 Stereoscopic imaging geometry provides with a very inefficient sampling of the lightfield.

On the contrary, the stereoscopic camera system performs a very inefficient sampling of the lightfield. The angular information is oversampled, while the spatial information is strongly undersampled (see Fig. 7). In this case the angular oversampling does not provide any advantage, while the spatial oversampling makes impossible to recover the original lightfield (see Fig. 8).

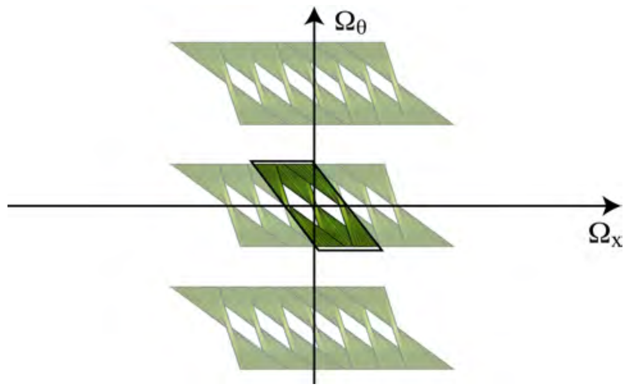

Fig. 8 The sampling obtained with stereoscopic pair produce a lightfield spectrum with strong overlapping.

Thus we can conclude that, although stereoscopic pairs have shown to be very useful for displaying of 3D movies or pictures big audiences, is very poor system when used for the tomography reconstruction of 3D scenes. Our opinion is that, given a fixed number of pixels. The optimum solution is an integral imaging system in which the number of views is equal to the number of pixels per view. The fulfillment of this condition permits the optimum, homogeneous sampling of the lightfield.

As the proof of the above statements we performed the experiment illustrated in the Fig. 9. In the experiment the distance between the camera lens to the doll and to the chart were set to 35 cm and 59 cm, respectively.

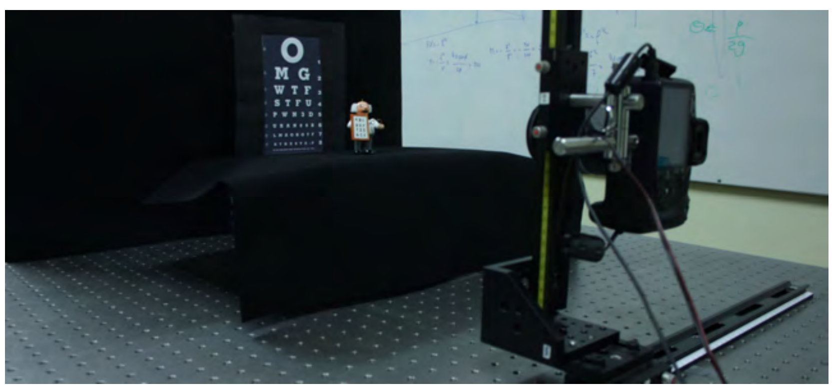

Fig. 9 Experimental setup.

With this setup [3] we captured first a set of 13x13 elemental images with 151x151 px each, obtained after displacing the camera in steps of 10 mm. Later, we captured 2 stereoscopic images with 1133x1701 px each. The distance between the two camera positions was set to 60 mm. In both cases the whole number of pixel was: 3.85 Mpx.

Next, in Fig. 10 we show a subset (13x10) of the collection of the captured elemental images. On the other hand, in Fig 11, we show the stereo pair captured with the digital camera. In both cases, the images are used for the calculation (by back-propagation algorithm) of different slices of original 3D scene.

In Fig 12 and Fig. 13 we show the images reconstructed at various depths. It is clear that although the resolution is better in the case of the stereo pair, the slicing capacity is almost null.

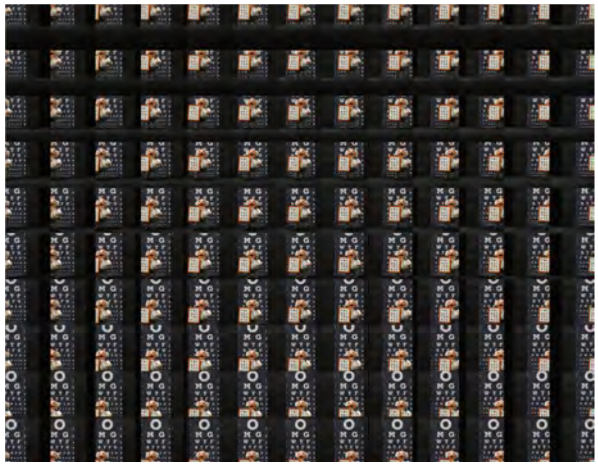

Fig. 10 The elemental images.

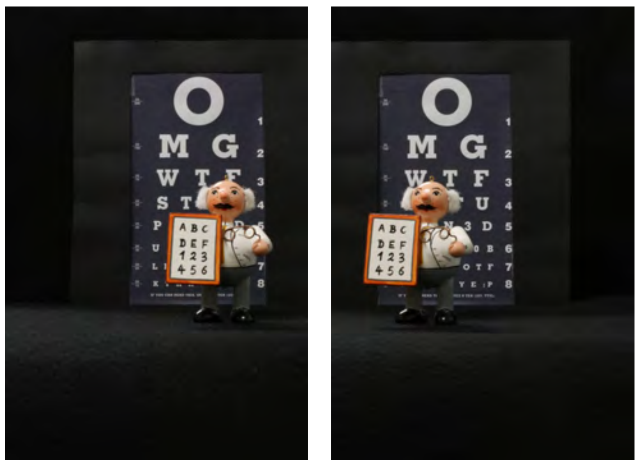

Fig. 11 The stereoscopic pair.

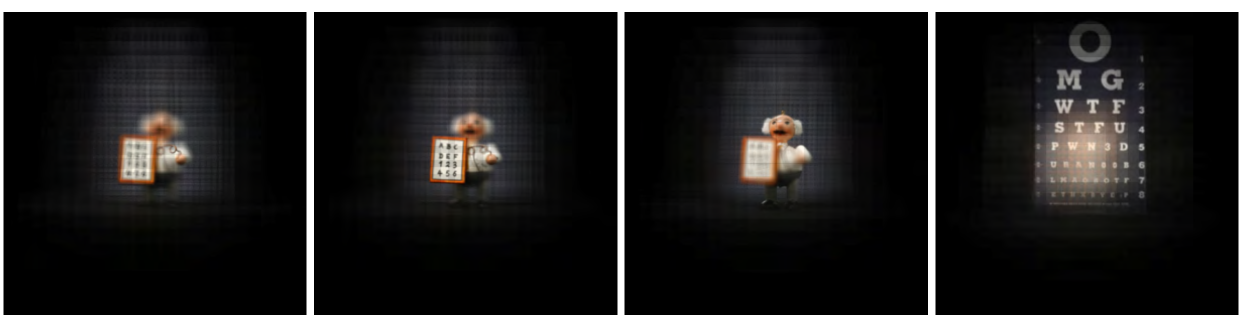

Fig. 12 Slices of the reconstructed 3D scene from the elemental images (from left to right): Spectacles; White chart ; Hair; Black chart.

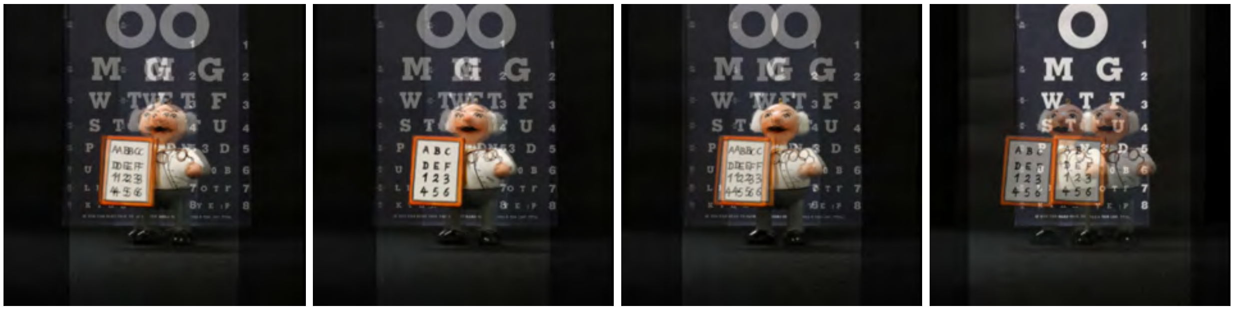

Fig. 13 Slices of the reconstructed 3D scene from the stereo pair (from left to right): Spectacles; White chart ; Hair; Black chart.

ACKNOWLEDGEMENTS

This work was supported in part by the Plan Nacional I+D+I under Grant FIS2009-9135, Ministerio de Ciencia e Innovación (Spain) and also by Generalitat Valenciana under Grant PROMETEO2009-077. Héctor Navarro gratefully acknowledges funding from the Generalitat Valencia (VALi+d predoctoral contract).

REFERENCES

[1] E. Adelson and J. Wang, "Single lens stereo with a plenoptic camera," IEEE Pattern Anal. Mach. Intell. 14, 99–106 (1992).

[2] Jin-Xiang Chai, Xin Tong, Shing-Chow Chan, and Heung-Yeung Shum, "Plenoptic sampling," *Proc. SIGGRAPH* 307–318 (2000).

[3] 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)

[4] J. W. Goodman, Introduction to Fourier Optics. New York: McGraw-Hill, 1996.

[5] Durand F., Holzschuch N., Soler C.,Chan E., Sillion F. X., "A frequency analysis of light transport," ACM Trans. Graph. 24, 3 (2005).