

IMAGE BASED RENDERING WITH DEPTH CAMERAS: HOW MANY ARE NEEDED?

Christopher Gilliam, James Pearson, Mike Brookes and Pier Luigi Dragotti

Electrical and Electronic Engineering Department, Imperial College London
Exhibition Road, SW7 2AZ, London, United Kingdom
phone: + (44) 2075946192, fax: + (44) 2075946234, email: p.dragotti@imperial.ac.uk
web: <http://www3.imperial.ac.uk/commssigproc>

ABSTRACT

Image based rendering (IBR) is a promising way to produce arbitrary views of a scene using images instead of object models. The emergence of low-price, fast, and reliable cameras for measuring depth will have a great impact on IBR in that depth measurements provide the perfect complementary information to the traditional color images. The issue then is to understand, given a certain scene of interest, how many depth images and how many color images are necessary in order to obtain good rendering results.

In this paper, we perform a spectral analysis of both multi-view depth images and multi-view color images in order to work out the relationship between the number of depth and color images needed. Our analysis is then validated using both synthetic and real images.

Index Terms— One, two, three, four, five

1. INTRODUCTION

Image Based Rendering (IBR) is a promising technique to render novel views from a set of available multi-view color images. Instead of rendering views of 3-D scenes by projecting objects and their textures, new views are rendered by interpolating available nearby images. The advantage of such a method is that it produces convincing photorealistic results since the interpolated viewpoints are obtained through combinations of real images. The main drawback is the fact that a huge amount of data needs to be captured.

Clearly, knowledge of the scene geometry reduces the number of images required. The interplay between geometry and sampling rate (number and spacing of cameras) has been extensively studied in the recent past (e.g., [1, 2, 3, 4, 5]). While the actual estimation of depth and geometry has been normally achieved using passive stereo with multiple cameras [1], James please include an all encompassing reference to depth estimation algorithms [2]. Unfortunately, 3-D reconstruction techniques from passive cameras, are still not reliable and do not work well in many cases. This fact has profoundly limited the use of IBR ideas. Recent advances in sensing technologies may soon allow large-scale deployment

of 3-D cameras using active depth sensing systems. These cameras are able to estimate depth and geometry with good accuracy and reliability, and for this reason can be very useful in IBR. The natural question then is to understand the interplay between the number of depth and colour cameras. Specifically, given a scene of interest with a certain geometry, how many depth cameras are necessary to infer the geometry and how many color cameras are then needed, given the inferred geometry, to render novel photorealistic view?

To answer this question we put ourselves in the typical Shannon sampling framework and perform a spectral analysis of both multi-view depth images and multi-view colour images. In that respect we continue the work of several researchers [1, 2, 3, 4]. In particular we use the formalism developed in a previous paper of ours [6] and expand it to include the case of depth cameras. We show that the interplay between the required number of depth and colour cameras mostly depends on the resolution of the colour cameras and the bandwidth of the texture of the scene. Our analysis is then validated using both synthetic and real images.

The paper is organized as follows:....

2. SPECTRAL ANALYSIS OF THE PLENOPTIC FUNCTION

2.1. Parameterization of the Plenoptic Function

At the heart of IBR is the idea that a scene can be represented as a collection of light rays emanating from the scene. As noted above, the complete parameterisation of the rays requires the seven dimensional plenoptic function. The number of dimensions, however, can be reduced by constraining the sensing setup. For example, the case when cameras lie on a plane leads to the 4-D lumigraph [7] or lightfield [8] parametrization. This parametrization is obtained by using two parallel planes: the camera plane (s, t) and the image plane (u, v) . The distance between the two planes is the focal length, f . Therefore, the function $p(s, t, u, v)$ represents the intensity of the light ray at camera location (s, t) and pixel location (u, v) (see Fig. 1). A further simplification, used in [1], is to fix s and u , corresponding to the situation where the camera posi-

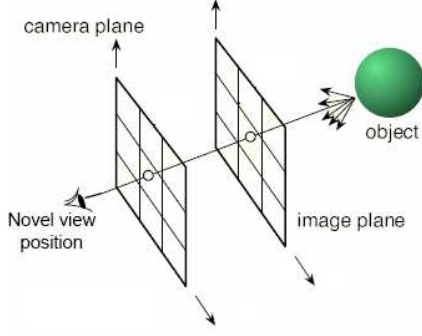


Fig. 1. Lightfield parameterization. A light ray is completely determined by its intersection with the camera and the image plane.

tions are constrained to a 1-D camera line and only one scan-line is considered in each image. In this case the light field is reduced to two dimensions: $p(t, v)$. Plotting this plenoptic representation in the (t, v) -space leads to the Epipolar Plane Image (EPI). It is interesting to note that in this case, a point in the scene is converted into a line in the plenoptic domain and lines with higher slope always occlude lines with smaller slope.

2.2. Bandwidth of the Plenoptic function

IBR can be seen as the problem of sampling and interpolating the plenoptic function. It is therefore natural to investigate the spectral properties of the plenoptic function in order to determine the correct Nyquist sampling rate and the correct reconstruction formula. Using the EPI parameterization the plenoptic spectrum is defined as $P(\omega_t, \omega_v) = \mathcal{F}\{p(t, v)\}$, where \mathcal{F} is the Fourier transform operator. The properties of the plenoptic spectrum were studied for the first time in [1]. By assuming a Lambertian scene with no occlusion, they showed that the spectrum is approximately bounded by lines related to the maximum and minimum depths of the scene and that finite camera resolution bandlimits the spectrum. This spectral analysis was extended in [3] to more general cases, in particular non-Lambertian and occluded scenes.

However, these analyses assume infinite scene width and cameras with infinite field of view when determining the plenoptic spectrum. In [6], we constrain our spectral analysis to cameras with FFoV and arrive at a close-form expression for the plenoptic spectrum of a slanted plane, of finite width, with bandlimited texture. An example of a slanted plane is illustrated in Fig. 2. The resulting plenoptic spectrum is band-unlimited in both ω_t and ω_v . However, by assuming the function is bandlimited to an essential bandwidth that contains 90% of the signal's energy, the EPI can be reconstructed up to a certain aliasing error. Under this assumption we determined a maximum acceptable spacing between cameras in

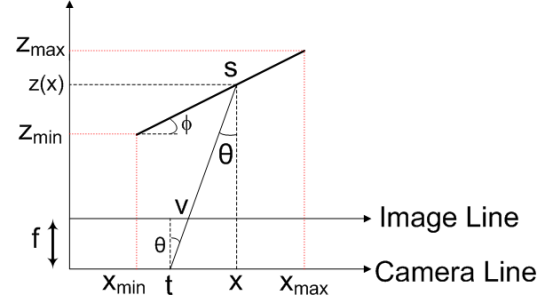


Fig. 2. Scene model of a slanted plane where z_{min} and z_{max} are the minimum and maximum depth, respectively, f is the focal length and h is the curvilinear coordinate. Note that θ is the viewing angle and ϕ is the slant of the plane.

order to reconstruct the plenoptic function.

Based on this spectral result, [?] formulates an algorithm that determines the optimum positioning for a finite number of cameras in order to sample a scene with a smoothly varying surface. However this algorithm requires as an input either the actual geometry or an estimate of it. **LINK TO DEPTH CAMERAS ETC**

3. THE DISPARITY FUNCTION

We propose treating the images obtained from the depth cameras as samples of a disparity function, $\delta(t, v)$, which describes the disparity of the scene captured at camera location t and pixel location v . Therefore, in a similar fashion to the plenoptic function, we perform spectral analysis on this disparity function with the aim of determining the minimum number of depth cameras required to reconstruct the scene geometry. This reconstructed geometry can then be fed into our adaptive plenoptic sampling algorithm.

In this paper we perform spectral analysis on the disparity function for a slanted plane. The scene geometry equations for a slanted plane are as follows

$$\mathcal{G}_s = \begin{cases} x = s \cos(\phi) + x_{min} \\ z(x) = (x - x_{min}) \tan(\phi) + z_{min} \end{cases} \quad (1)$$

where $x \in [x_{min}, x_{max}]$, $z \in [z_{min}, z_{max}]$ and ϕ is the angle between the plane and the line $z = z_{min}$. The finite width of the plane is

$$T = \frac{x_{max} - x_{min}}{\cos(\phi)} = \frac{z_{max} - z_{min}}{\sin(\phi)},$$

hence $s \in [0, T]$.

Having defined the scene geometry, we use the functional framework outlined in [2] in order to relate a point on the scene at $x, z(x)$ to a camera location t and pixel location v , hence

$$t = x - z(x) \frac{v}{f}. \quad (2)$$

Notice that by applying FFoV v is restricted to $v \in [-v_m, v_m]$. The relationship in (2) is restricted to a one to one mapping using a no-occlusion constraint

$$\frac{f}{v_m} > |z'(x)| = |\tan(\phi)|, \quad (3)$$

where $z'(x)$ is the first differential of $z(x)$ with respect to x . On a last note, the disparity of a point on the scene is constant regardless of the camera location, thus using (2) the following is true $\delta(t, v) = \hat{\delta}(x) \equiv 1/z(x)$.

3.1. Derivation of Disparity Spectrum

Starting with the Fourier transform of the disparity function under the FFoV and FSW constraints, we obtain

$$\begin{aligned} \Delta(\omega_t, \omega_v) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(t, v) e^{-j(\omega_t t + \omega_v v)} dt dv, \\ &\stackrel{(i)}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(1 - z'(x) \frac{v}{f}\right) \hat{\delta}(x) e^{-j(\omega_t(x - z(x)v/f) + \omega_v v)} dx dv, \\ &\stackrel{(ii)}{=} \int_{x_{min}}^{x_{max}} \hat{\delta}(x) e^{-j\omega_t x} \int_{-v_m}^{v_m} \left(1 - z'(x) \frac{v}{f}\right) e^{-j(\omega_v - z(x)\frac{\omega_t}{f})v} dv dx, \\ &\stackrel{(iii)}{=} 2v_m \int_{x_{min}}^{x_{max}} \hat{\delta}(x) \left[\text{sinc}(\omega_I) - j \frac{\tan(\phi)v_m}{f} \text{sinc}'(\omega_I) \right] e^{-j\omega_t x} dx, \end{aligned} \quad (4)$$

where $\text{sinc}'(\Omega_I)$ is the first derivative of the sinc function with respect to Ω_I , and $\omega_I = \omega_v v_m - (z(x)v_m \omega_t)/f$. Step (i) follows from applying (2) and assuming a Lambertian scene. Step (ii) follows from applying the FFoV and FSW constraints, and finally step (iii) from solving the integral in v .

By changing the variable of integration from x to s , using (1), the equation in (4) becomes

$$\begin{aligned} \Delta(\omega_t, \omega_v) &= 2v_m \left[\int_0^T \frac{\cos(\phi) \text{sinc}(\omega_s)}{s \sin(\phi) + z_{min}} e^{-j\omega_t \cos(\phi)s} ds \right. \\ &\quad \left. - j \frac{v_m \sin(\phi)}{f} \int_0^T \frac{\text{sinc}'(\omega_s)}{s \sin(\phi) + z_{min}} e^{-j\omega_t \cos(\phi)s} ds \right], \end{aligned} \quad (5)$$

where

$$\omega_s = \omega_v v_m - (s \sin(\phi) + z_{min}) \frac{v_m}{f} \omega_t.$$

Note that for clarity in this derivation, and without loss of generality, we set $x_{min} = 0$. At this point, we perform another change of variable, coupled with integration by parts, to obtain

$$\begin{aligned} \Delta(\omega_t, \omega_v) &= j \frac{2v_m^2}{f} \left(\frac{\text{sinc}(a)}{\omega_v v_m - a} e^{-j(a-b)c} - \frac{\text{sinc}(b)}{\omega_v v_m - b} \right. \\ &\quad \left. - e^{jbc} \int_b^a \frac{\text{sinc}(\omega_s)}{(\omega_v v_m - \omega_s)^2} e^{-j\omega_s c} d\omega_s \right) \end{aligned} \quad (6)$$

where

$$a = \omega_v v_m - \omega_t \frac{z_{max} v_m}{f}, \quad b = \omega_v v_m - \omega_t \frac{z_{min} v_m}{f},$$

$$\text{and} \quad c = \frac{-f}{\tan(\phi) v_m}.$$

Finally, the integral in (6) is solved by first splitting it into two, as follows

$$\frac{1}{2j} \left(\int_b^a \frac{(1 - e^{-jc_+ \omega_s}) d\omega_s}{\omega_s (\omega_v v_m - \omega_s)^2} - \int_b^a \frac{(1 - e^{-jc_- \omega_s}) d\omega_s}{\omega_s (\omega_v v_m - \omega_s)^2} \right),$$

where $c_+ = c + 1$ and $c_- = c - 1$. Then manipulating each of the integrals, using partial fractions, until they are defined in terms of a weighted combination of the identity [9]

$$\int_0^{jw} \frac{1 - e^{-\omega_s}}{\omega_s} d\omega_s = E_1(jw) + \ln(jw) + \gamma, \quad (7)$$

where $w \in \mathbb{R}$, $E_1(jw)$ is the exponential integral and γ is Euler's constant; for definitions see [9].

As a result the closed form expression for the disparity spectrum of a slanted plane is

$$\begin{aligned} \Delta(\omega_t, \omega_v) &= \frac{j2v_m}{f\omega_v} \left(\text{sinc}(a) e^{-j(a-b)c} - \text{sinc}(b) \right) + e^{j \frac{\omega_t z_{min}}{\tan(\phi)}} \\ &\quad \left(2v_m \ln \left(\frac{z_{max}}{z_{min}} \right) \left(\frac{\text{sinc}(\omega_v v_m)}{\tan(\phi)} - j \frac{v_m}{f} \text{sinc}'(\omega_v v_m) \right) \right. \\ &\quad + e^{-j\omega_v v_m} \left(\frac{1 + jc_+ \omega_v v_m}{f\omega_v^2} \right) \left[\zeta \{jw\} \right]_{w=-c_+ (\omega_v v_m - a)}^{w=-c_+ (\omega_v v_m - b)} \\ &\quad - e^{j\omega_v v_m} \left(\frac{1 + jc_- \omega_v v_m}{f\omega_v^2} \right) \left[\zeta \{jw\} \right]_{w=-c_- (\omega_v v_m - a)}^{w=-c_- (\omega_v v_m - b)} \\ &\quad \left. - \frac{1}{f\omega_v^2} \left(\left[\zeta \{jw\} \right]_{w=bc_+}^{w=ac_+} - \left[\zeta \{jw\} \right]_{w=bc_-}^{w=ac_-} \right) e^{jbc} \right) \end{aligned} \quad (8)$$

where

$$\zeta \{jw\} = \begin{cases} E_1(jw) + \ln(jw) + \gamma & \text{for } w \neq 0, \\ 0 & \text{for } w = 0. \end{cases} \quad (9)$$

Note that if $\omega_v = 0$ then

$$\begin{aligned} \Delta(\omega_t, 0) &= j2v_m \left(\frac{\text{sinc}(a)}{z_{max} \omega_t} e^{-j(a-b)c} - \frac{\text{sinc}(b)}{z_{min} \omega_t} \right) \\ &\quad + \frac{v_m^2}{f} e^{jbc} \left[\frac{j}{w} \sin(w) + \frac{c}{w} \sin(w) + \frac{j}{w} \cos(w) \right]_{w=b}^{w=a} \\ &\quad + \frac{v_m^2}{2f} e^{jbc} \left(c_+^2 \left[\zeta \{jw\} \right]_{w=bc_+}^{w=ac_+} - c_-^2 \left[\zeta \{jw\} \right]_{w=bc_-}^{w=ac_-} \right) \\ &\quad + \frac{2v_m}{\tan(\phi)} \ln \left(\frac{z_{max}}{z_{min}} \right) e^{jbc}, \end{aligned} \quad (10)$$

and, if $\omega_t = 0$ then

$$\Delta(0, \omega_v) = \frac{2v_m}{\tan(\phi)} \ln\left(\frac{z_{max}}{z_{min}}\right) \left(\text{sinc}(\omega_v v_m) - j \frac{v_m \tan(\phi)}{f} \text{sinc}'(\omega_v v_m) \right). \quad (11)$$

4. REFERENCES

- [1] J.X. Chai, X. Tong, S.C. Chan, and H.Y. Shum, “Plenoptic sampling,” in *Proc. of International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, July 2000, pp. 307–318.
- [2] M. Do, D. Marchand-Maillet, and M. Vetterli, “On the bandwidth of the plenoptic function,” *IEEE Trans. on Image Processing*, 2011, Preprint.
- [3] C. Zhang and T. Chen, “Spectral analysis for sampling image-based rendering data,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 11, pp. 1038–1050, November 2003.
- [4] C. Chen and D. Schonfeld, “Geometrical plenoptic sampling,” in *Proc. of International Conference on Image Processing (ICIP)*, Cairo (Egypt), November 2009.
- [5] C. Gilliam, P.L. Dragotti, and M. Brookes, “Adaptive plenoptic sampling,” in *Proc. of International Conference on Image Processing (ICIP)*, Brussels (Belgium), September 2011.
- [6] C. Gilliam, P.L. Dragotti, and M. Brookes, “A closed-form expression for the bandwidth of the plenoptic function under finite field of view constraints,” in *Proc. of International Conference on Image Processing (ICIP)*, Hong Kong, September 2010.
- [7] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, “The lumigraph,” in *Computer graphics (SIGGRAPH '96)*, 1996, pp. 43–54.
- [8] M. Levoy and P. Hanrahan, “Light field rendering,” in *Computer graphics (SIGGRAPH '96)*, 1996, pp. 31–42.
- [9] M. Abramowitz and I.A. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, 1964.