

# SAMPLING AND INTERPOLATION OF THE PLENOPTIC FUNCTION

*Amina Chebira*<sup>1</sup>

*Pier Luigi Dragotti*<sup>2</sup>

*Luciano Sbaiz*<sup>1</sup>

*Martin Vetterli*<sup>1,3</sup>

<sup>1</sup> Audio-Visual Communications Laboratory

Swiss Federal Institute of Technology, Lausanne, Switzerland.

<sup>2</sup> Electrical and Electronic Engineering Imperial College, London SW7-2BT, UK.

<sup>3</sup> EECS Dept., University of California at Berkeley, Berkeley CA 947020, USA

e-mail: {chebira,dragotti,sbaiz,vetterli}@lcavsun1.epfl.ch

## ABSTRACT

In this paper, we present reconstruction schemes for the plenoptic function. Using new sampling methods developed in [2, 4, 5], we show that, for some particular scenes, it is possible to perfectly reconstruct the plenoptic function from a finite number of cameras with finite resolution (Theorem 1 and Corollary 1). In more general cases, we demonstrate new ways of interpolating exactly the plenoptic function (Theorem 2). Finally, we show that it is possible to infer camera locations from a finite set of images. In all cases, we have perfect solutions due to the super-resolution property of the sampling. First numerical experiments on noisy observations show the potentiality of this new theoretical developments.

## 1. INTRODUCTION

Advances in device technology, networking and information processing have allowed the emergence of sensor networks. These are densely distributed sensor nodes with the capability to sense elements of the environment, make computations and communicate with other nodes or a central receiver. Signal sensors can be cheap temperature sensors, but also cameras or microphones.

In this new scenario, there are several problems to investigate that span from traditional signal processing problems (i.e., sampling, interpolation, compression and detection) to communication and information theory. Among them, of particular interest in this paper, is the sampling and interpolation problem.

In particular, we assume that our sensors are cheap cameras with low resolution and the question we want to address is how many cameras are needed to obtain a continuous rendering of a certain scene. That is, assume we have a certain number of cameras taking pictures of a scene from different positions, is there any sampling theorem that tells us the number of cameras necessary to reproduce that scene exactly? Or given a certain number of cameras can we interpolate other views?

The problem of rendering a 3D scene from a set of 2D images is known as image based rendering (IBR) and has attracted a lot of attention recently [8, 3, 9]. IBR is related to the concept of the plenoptic function introduced by Adelson and Bergen [1]. Assume that one is free to take photographs of a visual scene at any possible position, angle and time. Such a complete representation of that scene can be parameterized by a single function called the plenoptic function.

The sampling of the plenoptic function has been addressed by several researchers, and a first solution was given in [3]. For further results, see [9]. However, the sampling schemes developed in

those papers rely on the assumption that the plenoptic function is band-limited, which is usually not true. Therefore, in this case, the plenoptic function cannot be perfectly reconstructed from its set of samples.

Recently, it has been shown that it is possible to develop sampling schemes for classes of non-bandlimited signals [2, 4, 5]. A common features of these signals is that they have a parametric representation with a finite number of degrees of freedom.

In this paper, we develop new schemes for the sampling of the plenoptic function based on extension of the results in [2, 4, 5]. In particular, we show that it is possible to perfectly sample the plenoptic function of scenes with a finite number of degrees of freedom. Namely scenes made of points, planes, and other simple objects. Moreover, we show that, in other situations, we can exactly interpolate the plenoptic function with only a few cameras. We then propose to use these sampling schemes to solve a different problem, namely, to infer the cameras locations from a set of samples of the plenoptic function. Note that in all cases, we have perfect solutions, due to the super-resolution property of the sampling.

The paper is organized as follows. In the next section, after a brief review of some background material, we present our sampling and interpolating results (Theorem 1, Corollary 1 and Theorem 2). In Section 3, we address the problem of inferring the cameras locations from a set samples of the plenoptic function. Section 4 presents some numerical results. Finally, we conclude in Section 5.

## 2. INTERPOLATION AND SAMPLING THEOREMS FOR THE PLENOPTIC FUNCTION

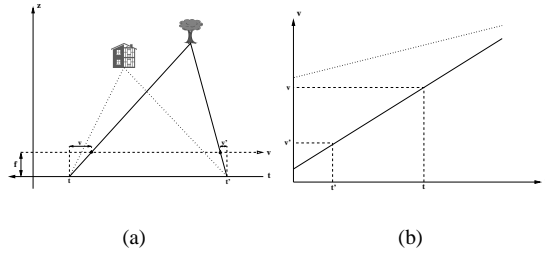
### 2.1. Background Material

The plenoptic function defined in [1] describes the intensity of each light ray in the world as a function of visual angle, wavelength, time and viewing position. Therefore, the plenoptic function is a 7D function

$$P_7 = P(\theta, \phi, \lambda, t, V_x, V_y, V_z). \quad (1)$$

By fixing the time (static scene) and the wavelength (fixed lighting conditions), this function becomes a 5D function. As stated in [6], we know that the 5D complete plenoptic function can be simplified to a 4D light field plenoptic function by restricting the scene (or conversely the camera view) to a bounding box or a convex hull:

$$P_4 = P(u, v, s, t), \quad (2)$$



**Fig. 1.** (a) Points of depth  $z(v, t)$  are observed by two cameras  $t'$  and  $t$  (b) 2D plenoptic functions of two points observed along the camera path

where  $(u, v)$  and  $(s, t)$  represent two parallel planes (the focal plane and the camera plane respectively) of the bounding box. Here we consider a pinhole camera model under which a point in space with coordinates  $\mathbf{X}=(x, y, z)^\top$  is mapped to the point  $(fx/z, fy/z, f)^\top$  on the focal plane defined by  $z = f$  where  $f$  is the focal length. By taking a  $t$ - $v$  slice of the 4D plenoptic function (which is equivalent to fixing the viewpoint) we reduce the plenoptic function to a 2D parametrization which is called 2D light field or epipolar plane image (EPI).

As shown in Figure 1, each point of an object is seen by two cameras lying on the  $t$ -axis. The  $v$ -axis corresponds to the focal line of the cameras. As in [3], we set up local coordinate on the focal line base on where the camera is. Each point is represented by a line in the  $t$ - $v$  plane where the slope of the line is related to the depth of the point through the following equation:

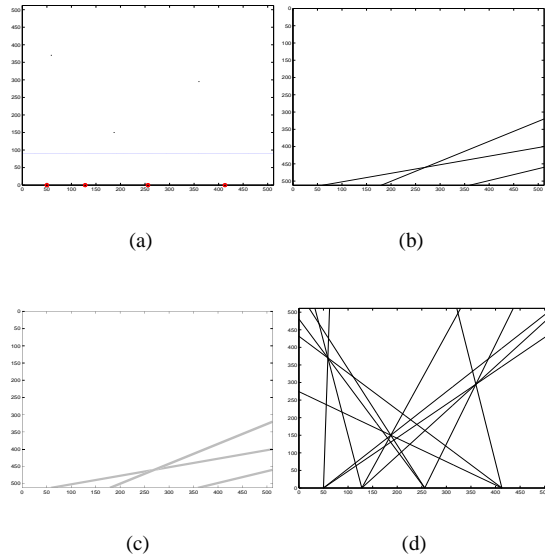
$$v - v' = f(t - t')/z. \quad (3)$$

Note that the image or focal line is placed in front of the camera centre. From now on, for simplicity, we will focus on the 2D plenoptic function. However, all results can be extended to the general case.

## 2.2. Interpolation And Sampling Results

The new sampling schemes we use [2] concern with a certain class of non-bandlimited signals namely those of finite rate of innovation such as streams of Diracs and piecewise polynomials. It has been shown that these signals can be sampled uniformly at (or above) the rate of innovation using either a sinc or a Gaussian kernel, and then be perfectly reconstructed. These results concern one-dimensional signals whereas [4] presents extensions to classes of 2D signals with finite complexity such as 2D Diracs (Poisson process), polygons and bilevel signals with piecewise polynomial boundaries. In [5], the case of the Radon transform of finite rate of innovation objects is studied, showing that a finite number of projections and a finite number of samples is sufficient for an exact reconstruction of certain simple objects such as Diracs.

In this work, we extend these results to the case of the plenoptic function. We show that it is possible to have an exact reconstruction of the plenoptic function of simple objects from a finite number of samples and a finite number of cameras. That is, assume a finite number of cameras with finite resolution and some lens distortion, then the plenoptic function can be recovered from this finite number of blurred and sampled images.



**Fig. 2.** Plenoptic function of a stream of Diracs and reconstruction (a) Original scene made up of  $M = 3$  Diracs, (b) 2D plenoptic functions (c) blurred plenoptic functions, (d) reconstruction of the points from the projecting lines. The Diracs correspond to points where exactly  $M + 1 = 4$  lines intersect.

Consider a set of  $M$  singular points (2D Diracs generated by a Poisson process) and a set of  $K$  cameras. Let  $S_i(v)$ ,  $i = 0, 1, \dots, K - 1$  be their perspective projections of optical centres  $t_i$ .  $S_i(v)$  can be represented as a weighted sum of at most  $M$  1-D Diracs lying on the image line. In practice, we observe a sampled and distorted version of these projections, but as stated in [2], the signal  $S_i(v)$  made up of maximum  $M$  Diracs can be recovered from a set of  $2M$  samples by using either the sinc or the Gaussian sampling kernel. Therefore, the perspective projections can be reconstructed. Moreover, one can prove (see [5]) that it is possible to recover the original scene from  $K \geq M + 1$  projecting lines. Finally, the weights of the singular points can be found by solving a linear system of equations. Hence, we can state the following theorem:

**Theorem 1:** *Consider a scene made up of  $M$  singular points (Diracs) and let its perspective projection be convolved with a sinc or a Gaussian sampling kernel. By taking  $K \geq M + 1$  perspective projections and for each projection, at least  $2M$  samples, we can perfectly reconstruct the scene.*

Figure 2 shows an example of Theorem 1 with  $M = 3$  and four cameras ( $K = 4$ ). Theorem 1, basically, states that with  $K \geq M + 1$  cameras (or with  $K$  different pictures of the same scene), we can recover exactly a visual scene made of  $M$  Diracs. In many cases, however, it is possible to have the same result with only two cameras. In fact, in IBR, one has access to the perspective projections and can label the points (Diracs) in these projections. Then, in this case, and in absence of occlusions, two cameras are sufficient to reconstruct the plenoptic function of a scene made up of an arbitrary but finite number of Diracs. For the sake of clarity, we state this in the following corollary:

**Corollary 1:** Consider a scene made up of  $M$  labelled Diracs and two perspective projections of this scene with no occlusions. Then,  $N \geq 2M$  sample values of each projection taken with a sinc or Gaussian sampling kernel are a sufficient representation of the scene.

This result can be extended to the class of piecewise constant signals. Assume, the scene is made up of surfaces of different uniform color. In this case, the perspective projection on the image plane is a piecewise constant signal containing  $L$  pieces. The positions  $t_0$  and  $t_1$  of the cameras are such that none of these pieces is occluded in the range  $[t_0, t_1]$ . Using [2], we also know that the two views given by the cameras can be recovered from  $N \geq 2L + 1$  samples. From the reconstructed views and by means of projective geometry, we can then reconstruct any view in between  $[t_0, t_1]$ . Thus we can derive the following result:

**Theorem 2:** Consider a scene made up of surfaces containing  $L$  pieces each of different uniform color. Two cameras are at the positions  $t_0$  and  $t_1$ . Suppose that these positions are such that none of the pieces is occluded in the range  $[t_0, t_1]$ . Then, by using a sinc sampling kernel and at least  $2L + 1$  samples of each view taken by the cameras, we can:

1. recover perfectly the two views.
2. interpolate the plenoptic function in between the two views.

### 3. SUPER-RESOLUTION CAMERA LOCATION

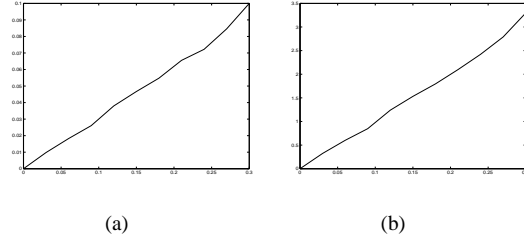
Until now, we have considered the case where cameras locations are known. What if we do not know this information? In this case, what additional knowledge do we need in order to completely recover cameras locations and hence the scene? The problem of finding the camera positions and orientations has been extensively studied in computer vision (see for instance [7]).

Our novel approach to this problem is to use the sampling schemes for signals of finite rate of innovation to find exact cameras locations. These methods lead to perfect reconstruction from a finite number of samples of the perspective projections and allows us to recover the exact positions of the cameras. In the general case, the system of equations we have to solve is polynomial but if we suppose that the cameras have the same orientation with then the system becomes linear and in this case we can derive the following theorem:

**Theorem 3:** Consider a set of  $N$  aligned cameras looking at a scene composed of  $N$  singular points. Suppose that we know the exact location of one of these points (reference point). Then, we can perfectly recover the  $N$  cameras positions and the original scene by taking at least  $2(N - 1)$  samples for each view taken by a camera.

To prove this theorem, we just have to notice that the coordinates of the reference point are used to find the locations of the  $N$  cameras and then we use our algorithm and the positions of the cameras to reconstruct the locations of the  $N - 1$  points. We can illustrate this by considering the case where  $N = 2$ : let  $(x_1, z_1)$  and  $(x_2, z_2)$  be the coordinates of two points  $P_1$  and  $P_2$ , respectively and let  $(x_{cj}, z_{cj}, \theta_j)$  be the coordinates of the camera  $j$  ( $1 \leq j \leq N$ ) where  $z_{cj}$  is the same for each camera and is known (see Figure 1(a)). Let  $P_1$  be the reference point. Then, the system of  $N$  equations we have to solve is of the form:

$$v_{j1} = \frac{(x_{cj} - x_1) + \tan \theta_j (z_{cj} - z_1)}{\tan \theta_j (x_{cj} - x_1) - (z_{cj} - z_1)} f \quad (4)$$



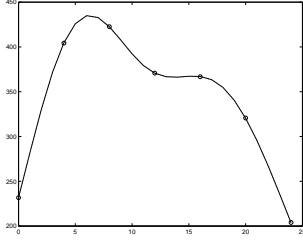
**Fig. 3.** Monte Carlo simulations for the estimation of Diracs positions (a) and weight (b). Horizontal axis is the additive noise standard deviation and vertical axis the estimation error standard deviation.

where  $v_{j1}$  is the image of the reference point on the camera  $j$ . If  $\theta = 0$ , then the system is linear. The variables  $v_{j1}$ ,  $j = 1, \dots, N$  are observed but thanks to the sampling theorem given in [2], we can know them exactly (i.e with infinite precision) and thus we can solve the system of equations to find the exact values of  $x_{cj}$ .

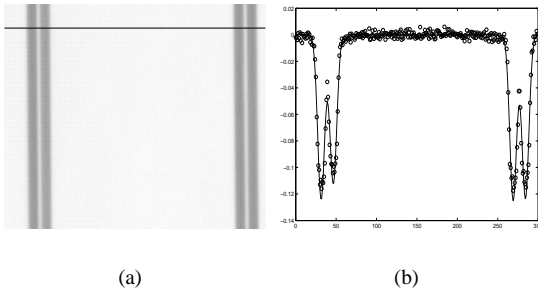
### 4. SIMULATION RESULTS

In Section 2, it has been shown that the positions of  $M$  singular points can be reconstructed perfectly taking  $N$  samples on  $K$  projections. The reconstruction is obtained in two steps: first the  $N$  samples are used to find the position of the singular points on each projection, then such positions are used to reconstruct the positions on the  $t-v$  plane. In a practical setup, the precision of the method is limited by the sensors, which introduce distortion and noise that limit the precision of the first step of the reconstruction algorithm. In order to investigate the robustness of the algorithm with respect to noise, Monte Carlo simulations have been performed. Namely, a noiseless signal made up of  $M = 3$  weighted Diracs has been generated. Diracs positions were at  $v_0 = 5.3$ ,  $v_1 = 16.2$ ,  $v_2 = 22.7$  and weights  $a_0 = 404$ ,  $a_1 = 275$ ,  $a_2 = 126$ . The signal has been filtered with the Gaussian kernel with parameter  $\sigma_K = 5$  and  $N = 7$  samples has been considered uniformly spaced between  $v = 0$  and  $v = 24$ . Noisy samples have been obtained adding Gaussian noise and the position and weights of the Diracs reconstructed with the algorithm presented in [2]. Different level of noise standard deviation  $\sigma_n$  has been considered and 200 experiments have been run for each noise level. One of the experiments corresponding to  $\sigma_n = 0.3$  is depicted in Figure 4. The standard deviation of the parameter estimation error is illustrated in Figure 3 for the Diracs positions and weights respectively. It should be pointed out that, in the current implementation of the algorithm, the parameter  $\sigma_K$  cannot be too small to avoid ill conditioning of the system. Such a parameter influences also the relation between the noise standard deviation,  $\sigma_n$  and the estimation error. In fact, for a fixed value  $\sigma_n$ , a bigger  $\sigma_K$  increases the number of meaningful samples, reducing the estimation error.

To be able to use this type of algorithm on a real setup, one must be careful in avoiding potential additional sources of errors, such as lens distortion, poor precision of lens calibration, sensor noise and non linearities and bad illumination conditions. To make a preliminary evaluation of the proposed method we determined the estimation errors indirectly, using a pattern containing four ver-



**Fig. 4.** One experiment of the simulation with a noise standard deviation  $\sigma_n = 0.3$ . Points correspond to the  $N = 7$  samples used by the algorithm. The line correspond to the reconstructed values and is practically superposed to the ground truth values.



**Fig. 5.** An image of the pattern used to measure the estimation error on diracs positions (a). The dots represent the sensor data corresponding to the marked line of the pattern. The solid line corresponds to the reconstructed stream of Diracs (b).

tical lines at positions  $v_0, v_1, v_2$  and  $v_3$ . The positions are chosen so that  $v_1 - v_0 = v_3 - v_2 \gg v_2 - v_0$ . The four lines of the pattern are printed on a transparent slide placed in front of a uniform light source. The images of the pattern were acquired using a digital camera with the settings chosen to operate on the linear region of the CCD. The focus control was deliberately set to have a slightly defocused image, to reduce the influence of print imprecision and light conditions. Also, a defocused image corresponds approximately to increase the parameter  $\sigma_K$  of the kernel, avoiding the problems of ill-conditioning mentioned above.

To avoid the influence of lens distortion, only the center of the image was used, namely a rectangle of 100x300 pixels (the image size was 1456x2224). An example of the acquired image is given in Figure 5(a) and the samples for the marked line are reported in Figure 5(b). Since the thickness of the lines is small with respect to the parameter  $\sigma_K$  of the Gaussian kernel, every line can be modelled as a Dirac impulse in the horizontal direction and the algorithm in [2] applied with  $K = 4$ . The estimations of the positions of the Diracs,  $\hat{v}_0, \hat{v}_1, \hat{v}_2$  and  $\hat{v}_3$  can be used to estimate the ratio:

$$r = \frac{v_1 - v_0}{v_2 - v_0}. \quad (5)$$

The advantage of using a ratio between distances rather than computing an absolute measure on the pattern is that knowledge of lens parameters (such as focal length and position of the nodal point) is not required, avoiding a potential source of errors. It can be easily

verified that the estimation errors  $e_0 = \hat{v}_0 - v_0, e_1 = \hat{v}_1 - v_1$  on the Diracs' positions and the ratio estimation error are related by

$$e_1 - e_0 \simeq (\hat{r} - r)(v_2 - v_0). \quad (6)$$

If we model  $e_0, e_1$  as i.i.d. variables, one can compute the estimation error standard deviation  $\sigma_e \simeq (\hat{r} - r)(v_2 - v_0)/\sqrt{2}$ . This method has been used to evaluate the estimation error of the proposed technique, obtaining  $\sigma_e \simeq 0.2$  pixels. That is, on this blurred and noisy image case, we have obtained a precision of about 1/5 of a pixel, which is a promising preliminary result.

## 5. CONCLUSIONS

We have developed perfect reconstruction methods for simple scene models using a small number of cameras. These methods are based on new sampling schemes for signals of finite rate of innovation [2, 5]. We have derived interesting results for a perfect interpolation of the plenoptic function of piecewise linear objects. We also have shown that the cameras locations can be recovered if we know the position of one point in the scene. Finally, preliminary results on real cameras have shown the potentiality of these new sampling schemes.

## 6. REFERENCES

- [1] E.H. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," In Computational Models of Visual Processing, pages 3-20. MIT Press, 1991.
- [2] M. Vetterli, P. Marziliano, T. Blu, "Sampling signals with finite rate of innovation," IEEE Trans. Signal Processing, vol. 50, pages 1417-1428, June 2002.
- [3] J.X. Chai, X. Tong, S.C. Chan, H.Y. Shum, "Plenoptic sampling," In Proc. SIGGRAPH, 2000.
- [4] I.Maravic, M. Vetterli, "Sampling results for classes of non-bandlimited 2-D signals," IEEE Trans. Signal Processing, to appear 2003.
- [5] I.Maravic, M. Vetterli, "A sampling theorem for the Radon transform of finite complexity objects," Int. Conference on Acoustics, Speech and Signal Processing, May 2002.
- [6] M. Levoy, P.Hanrahan, "Light field rendering," In Computer Graphics Proc, Annual Conf. Series, pages 31-42, Proc. SIGGRAPH, 1996.
- [7] R. Hartley, A. Zisserman, "Multiple view geometry in computer vision," Cambridge University Press, 2001.
- [8] S. J. Gortler, R. Grzeszczuk, R. Szeliski, M. F. Cohen, "The lumigraph," In Computer Graphics Proc, Annual Conf. Series, pages 43-54, Proc. SIGGRAPH, 1996.
- [9] C. Zhang, T. Chen, "Generalized plenoptic sampling," Technical Report, Advanced Multimedia Processing Laboratory, Carnegie Mellon University, September 2001.