

GRADIENT-BASED DEPTH ESTIMATION FROM 4D LIGHT FIELDS

Don Dansereau, Len Bruton

Dept. of Electrical and Computer Engineering,
University of Calgary, Alberta, Canada

ABSTRACT

It is shown that an infinitesimally small surface element of a Lambertian scene exists as a plane of constant value in a 4D light field, where the orientation of the plane is determined by the depth of the element in the scene. By applying 2D gradient operators to appropriate subsets of the light field, the orientations of these constant-valued planes, and thus the depths of the corresponding elements of the scene, may be estimated. The redundancy associated with using three color channels, and having two depth estimates based on orthogonal 2D gradient estimates, is resolved using a weighted sum based on the confidence of each estimate.

1. INTRODUCTION

Image-based rendering has gained attention as a fast alternative to geometric model-based rendering. Light field rendering [1] and the Lumigraph [2] are two similar image-based rendering techniques which seek to model a 4D subset of the more general 7D plenoptic function [3] associated with a scene. In this way, the set of light rays permeating a scene is represented, rather than the geometry of the objects within the scene.

The 7D plenoptic function describes the light rays in a scene as a function of position, orientation, spectral content, and time. This can be simplified to a 4D function [1] by considering only the value of each ray as a function of its position and orientation in a static scene, and by constraining each ray to have the same value at every point along its direction of propagation. This disallows scenes in which the medium attenuates light as it propagates, and it fails to completely model the behavior of rays as they are occluded. These limitations are not an issue for scenes in a clear medium such as air, and for which the camera is not allowed to move behind occluding objects.

The 4D light field typically parameterizes light rays using the two-plane parameterization (2PP), as depicted in Fig. 1. Each ray is described by its point of intersection with two reference planes: the s, t plane given by $z = 0$, and the u, v plane which is parallel to the s, t plane at some positive separation $z = d$. A full light field may consist of multiple sets of such planes, though this paper will deal only with a single set of reference planes. Note also that each sample of a light field can be taken as a grayscale intensity, though extension to utilize color samples given as red, green and blue components is a simple matter of keeping one light field for each color channel, and repeating each operation accordingly, as is done in most image processing applications.

Because light fields accurately model scenes which are geometrically complex, they are well suited to act as an intermediary between the real world and computer vision algorithms. By storing

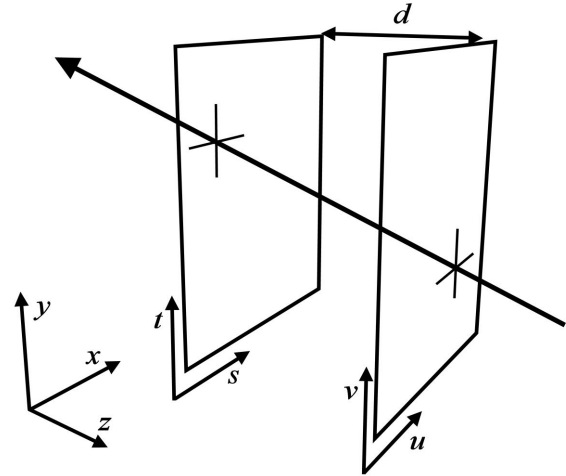


Fig. 1. Two-plane parameterization of light rays.

a large amount of information about a scene prior to processing, light fields allow more informed decision making, and for simple algorithms to accomplish complex tasks.

This paper is focused on the task of estimating the shape of a scene modeled using a light field. Normally shape estimation is quite involved, with complexity proportional to that of the scene geometry. Because a light field model of the scene is utilized, simple methods suffice, and the speed and accuracy of these methods are independent of scene complexity. Because of their simplicity, the techniques described are extremely robust, and may have applications as diverse as robot navigation [4] and object recognition [5], for example.

Throughout this paper, the continuous-domain light field will be denoted as $L_{cont}(s, t, u, v)$, and the discrete-domain version as $L(n_s, n_t, n_u, n_v)$, where \mathbf{n} is the discrete-domain index of the signal. The technique described here will assume equal sample rates in the four dimensions – extension to differing sample rates requires appropriate adjustment.

2. THE POINT-PLANE CORRESPONDENCE

2.1. The Omni-Directional Point Light Source

Fig. 2 depicts a 2D slice along s and u of a subset of the rays emanating from an omni-directional point source of light at the position $\mathbf{P} = [P_x, P_y, P_z]$. It is clear from this figure that for any given point on the s, t plane there is only one point on the u, v plane for which a ray will intersect the light source. The re-

This work was supported in part by funding from NSERC.

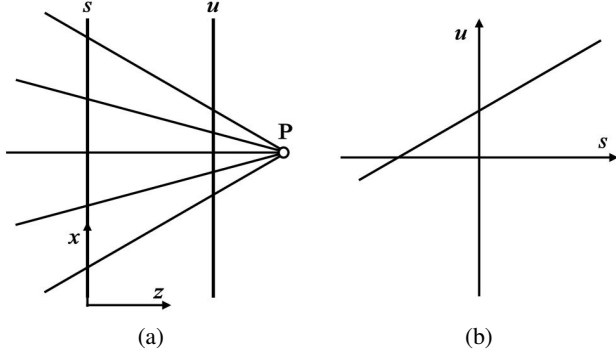


Fig. 2. The point-plane correspondence: a) 2D slice of a point source of light shown with the two reference planes; b) 2D slice of the corresponding light field.

sult is that an s, u slice of the corresponding continuous-domain light field $L_{cont}(s, t, u, v)$ takes the form of a line, as depicted in Fig. 2b). The equation of this line is given by

$$(d/P_z - 1)s + u = P_x d/P_z, \quad (1)$$

where d is the separation of the reference planes. The behavior in the t and v dimensions is similar, and can be expressed as

$$(d/P_z - 1)t + v = P_y d/P_z. \quad (2)$$

In 4D space, (1) and (2) are the equations of two hyperplanes with normals in the directions $\mathbf{D}_1 = [d/P_z - 1, 0, 1, 0]$ and $\mathbf{D}_2 = [0, d/P_z - 1, 0, 1]$, respectively [6]. The set of points in the light field which satisfy both (1) and (2) belongs to a plane defined by the intersection of these two hyperplanes. Points belonging to this plane of intersection correspond to rays emanating from the point light source and so take on the value of that light source, while all other points in the light field have a value of zero. Thus, an omni-directional point light source is a plane of constant value in the light field, where this plane is the solution of (1) and (2).

2.2. A Lambertian Surface

A Lambertian surface is one with ideal diffuse reflectance – that is, the luminance for any given point on such a surface is independent of viewing angle. In this sense, an infinitesimally small area of a Lambertian surface behaves similarly to an omni-directional point light source, and will therefore exist in a light field as a plane of constant value according to the point-plane correspondence. A complex scene may be modeled as a collection of such infinitesimally small surface elements, and thus as the superposition of many constant-valued planes in a light field. The orientation of each plane is determined by the depth of the corresponding surface element. Note that this superposition assumes the absence of occlusions – occluded regions will appear as truncated planes in a light field model. These observations can be seen in Fig. 3, which is a slice in s, u of a light field of a Lambertian scene containing elements at different depths.

Because the orientations of the planes that make up a light field depend only on the depths of the corresponding surface elements, the depth of each surface element may be estimated by finding the orientation of the corresponding plane.

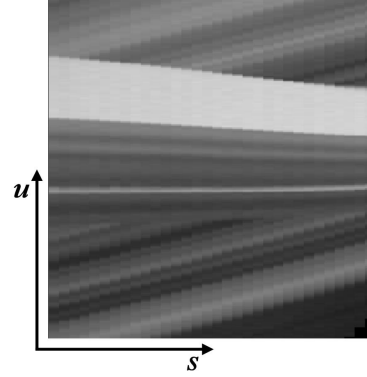


Fig. 3. A slice in s and u of a simple light field.

3. DEPTH ESTIMATION

The technique we propose is to estimate the orientation of the plane passing through each light field sample, and use this orientation to estimate the depth of the corresponding scene element. Because a separate depth estimate is formed for each light field sample, occluded regions are treated correctly.

3.1. Plane Orientation Estimation

The orientation of a plane is difficult to estimate in 4D. In the general case, two 4D vectors are required to do so. Thankfully, the planes in a light field are constrained to have the same orientation in the s, u directions as in the t, v directions. This observation will allow the use of 2D gradient operators, applied in slices in s, u and t, v , in estimating a plane's orientation, and it will introduce redundancy which can be used to validate the results. The use of gradients to estimate depth is mentioned briefly in [7].

Observing an s, u slice of a simple light field, such as the one shown in Fig. 3, it is clear that a 2D gradient operator, applied at some point in the slice, will yield a gradient vector which points orthogonal to the plane passing through that point. The basic approach associated with gradient-based depth estimation, then, is to use a 2D gradient operator in the s and u dimensions to estimate the orientation of the plane passing through each light field sample. Gradient operators applied in t and v should yield the same results, introducing a redundancy of a factor of two. In fact, because the 2D gradient operator can be applied to each color channel of the light field independently, a total redundancy of a factor of six is introduced.

The 2D gradient operator, applied to a single color channel in the s and u directions, is defined as

$$\nabla^{su} L(\mathbf{n}) = \left[\frac{\partial L(\mathbf{n})}{\partial s}, \frac{\partial L(\mathbf{n})}{\partial u} \right]. \quad (3)$$

The values of the partial derivatives can be estimated using a number of methods. One of the simpler methods, which has a relatively high immunity to noise, utilizes 2D convolution in s and u , as in

$$\frac{\partial L(\mathbf{n})}{\partial s} \simeq L(n_s, n_u) * \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}. \quad (4)$$

Extension to the partial derivative in u , and of both the gradient operator and partial derivatives to operate on slices in t and v , is trivial.

3.2. Estimating Depth

Given the gradient vector, the slope of the plane passing through each sample in the light field can be found, and from this slope the depth of the corresponding point in the scene can be found. From the point-plane correspondence, the slope of the plane corresponding to a point in the scene is given, in the s and u directions, by

$$m_{plane} = 1 - \frac{d}{P_z}. \quad (5)$$

The direction of the gradient vector can also be expressed as a slope, as in

$$m_{\nabla} = \frac{\nabla_u^{su} L(\mathbf{n})}{\nabla_s^{su} L(\mathbf{n})} = \frac{\partial L(\mathbf{n})/\partial u}{\partial L(\mathbf{n})/\partial s}. \quad (6)$$

Assuming no regions of constant value, for which the magnitude of the gradient vector is zero, the gradient vector will always point orthogonal to the plane. This means that the relationship between the slope of the plane and the slope of the gradient vector can be expressed as $m_{plane} = -m_{\nabla}^{-1}$. Rearranging to solve for P_z yields the equation

$$P_z = \frac{d}{1 + \frac{\partial L(\mathbf{n})/\partial s}{\partial L(\mathbf{n})/\partial u}}, \quad (7)$$

which is easily generalized to the t and v dimensions. By applying this equation throughout the light field, the shape of the scene that it models is estimated.

3.3. Consolidating Redundant Results

Because of the redundancy associated with having three independent color channels, with two independent depth estimates per channel, some method of optimally combining the estimates is in order. One way of doing this is to take the weighted sum of the six depth estimates, where the weight is taken as some measure of confidence. Given that the magnitude of the gradient vector, $\|\nabla L(\mathbf{n})\|$, is essentially an indication of the contrast of the light field at each sample, it is a good indicator of confidence. Areas of low contrast, which yield little information about a scene's shape, will have short gradient vectors, while areas of high contrast will have long gradient vectors. The weighted sum can be expressed as

$$\bar{P}_z = \frac{\sum_{i=0}^5 P_{z,i} \|\nabla_i L(\mathbf{n})\|}{\sum_{i=0}^5 \|\nabla_i L(\mathbf{n})\|}, \quad (8)$$

where the i^{th} gradient and depth estimate correspond to one of the six unique combinations of color channel and direction. Because the denominator of this expression is an indication of overall confidence, it can be used in a thresholding operation – allowing depth estimates with inadequate support to be ignored altogether.

3.4. Obtaining a Dense Estimate

When using the denominator of (8) as the basis for a thresholding operation, the output is sparse for scenes with regions of constant value. Some method of filling the gaps in the estimate is desirable. This can be achieved using a simple region growing algorithm. By iteratively assigning each empty estimate the average value of all of its nearest neighbors along s , t , u and v , the gaps can be filled.

A further improvement in the output can be obtained through the use of a 4D lowpass filter – a simple 4D moving average filter is sufficient – to increase immunity to noise, aliasing, occlusions and specular reflections. The use of a lowpass filter is particularly appropriate, as it goes some way towards consolidating the large amount of information associated with having a depth estimate for every light field sample.

4. RESULTS

Gradient-based depth estimation was applied to a light field with geometric parameters as summarized in Table 1. The light field was measured using a gantry system and contains occlusions. A rendered view of the input light field is shown in Fig. 4. The background is a poster of a supernova imaged by the Dominion Radio Astrophysical Observatory in British Columbia, Canada. The foreground is a beer coaster mounted at approximately 45 degrees to the reference planes – a wooden dowel can be seen holding the coaster in place. The poster is at a depth of 66 cm, and the coaster occupies a range of depths from about 40 cm to 50 cm. With this size of light field, gradient-based depth estimation takes about 35 ms to form depth estimates in a single u , v slice of the light field, and about 35 s to fill the entire light field structure.

Table 1. Input light field parameters.

Color channels	3
s, t samples	32
u, v samples	128
s, t size (cm)	21
u, v size (cm)	15
Separation d (cm)	45

The results of the depth estimation are shown in Fig. 5. This figure visualizes the depth estimates as intensity – pure black represents a depth of 30 cm, while pure white represents a depth of 70 cm. The poor estimates associated with areas of constant value are clear in a). These regions were reduced by applying the thresholding technique, with a threshold of 1.0 – that is, the average magnitude of the gradient vectors had to exceed 1.0 before the corresponding depth estimate was used. The results of the thresholding are shown in b) – regions that have been ignored due to thresholding appear black. Clearly the thresholded estimate is more desirable, having the advantage of allowing regions of low confidence to be ignored.

Applying a region growing algorithm, followed by a simple 4D moving average filter, yields the result shown in Fig. 6 – in this case, a window size of 4 samples was used. These techniques have clearly yielded a highly accurate, dense depth estimate.

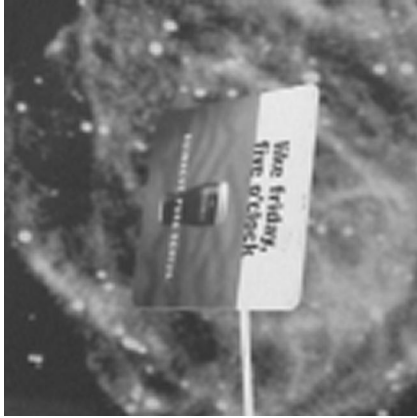


Fig. 4. Input light field.

5. CONCLUSIONS

A technique for estimating the depth of each visible point in a scene modeled as a light field was described. The technique utilizes 2D gradient operators applied to slices of the light field in s, u and in t, v . The estimates are redundant by a factor of six because of the independent estimates formed in s, u and t, v , as well as in the three color channels. This redundancy was consolidated by weighting each estimate with a corresponding confidence, where confidence was taken as the magnitude of the corresponding gradient vector.

Gradient-based depth estimation was shown to be effective with a light field containing occlusions. The results were improved by ignoring estimates for which the average confidence did not exceed a given threshold. Regions in which no estimates were formed were filled using a simple region-growing algorithm, and smoothed using a 4D moving average filter.

Though they first came about in the context of synthesizing graphics, image-based techniques have a significant potential in the field of scene analysis. Specific applications of the depth estimation technique presented here might include robot navigation [4], scene modeling [8], and object or face recognition [5] [9]. The possibility of applying simple and robust techniques to accomplish complex tasks is exciting.

Future work might involve the optimization of the technique

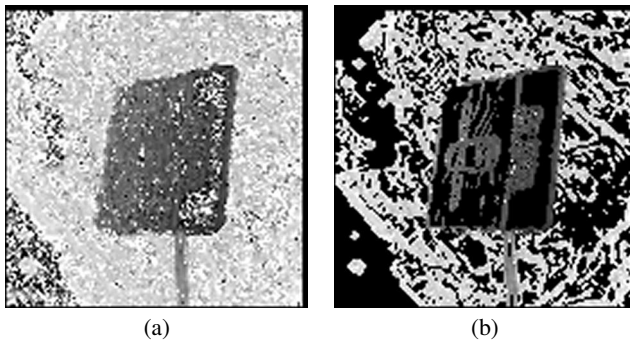


Fig. 5. Results of gradient-based depth estimation a) without thresholding, and b) with thresholding.

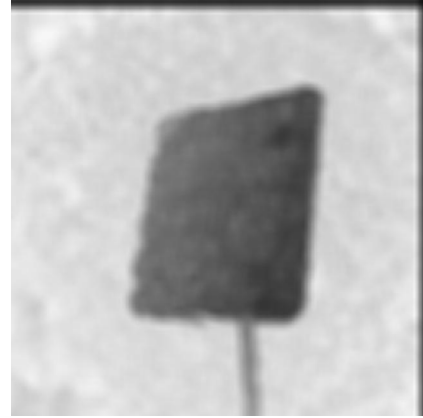


Fig. 6. Results of applying region growing and a 4D moving average filter to the output of gradient-based depth estimation.

described in this paper, through the use of parallel processing, for example, to operate in real-time. Optimization for operation in the presence of occlusions and specular reflections may also prove interesting.

6. REFERENCES

- [1] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of SIGGRAPH*, 1996, pp. 31–42.
- [2] S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The lumigraph," in *Proceedings of SIGGRAPH*, August 1996, pp. 31–42.
- [3] E.H. Adelson and J.R. Bergen, "The plenoptic function and the elements of early vision," *Computational Models of Visual Processing*, pp. 3–20, 1991.
- [4] B. Heigl, J. Denzler, and H. Niemann, "Combining computer graphics and computer vision for probabilistic visual robot navigation," in *Proceedings of SPIE's 14th Annual International Symposium on Aerospace/Defense Sensing, Simulation, and Controls*, Orlando, Florida, April 2000.
- [5] B. Heigl, J. Denzler, and H. Niemann, "On the application of light field reconstruction for statistical object recognition," in *European Signal Processing Conference (EUSIPCO)*, 1998, pp. 1101–1105.
- [6] S. Lang, *Calculus of Several Variables*, p. 38, Springer Verlag, 3rd edition, January 1995.
- [7] S. Baker, T. Sim, and T. Kanade, "When is the shape of a scene unique given its light-field: A fundamental theorem of 3D vision?," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 100–109, January 2003.
- [8] C. Vogelgsang, B. Heigl, Greiner G., and H. Niemann, "Automatic image-based scene model acquisition and visualization," in *Workshop Vision, Modeling and Visualization*, Saarbrücken, Germany, November 2000, pp. 189–198.
- [9] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields," Technical report, Robotics Institute, Carnegie Mellon University, CMU-RI-TR-02-20, August 2002.