DEPTH DETECTION OF LIGHT FIELD

Yi-Hao Kao, Chia-Kai Liang, Li-Wen Chang, and Homer H. Chen

National Taiwan University
1, Sec. 4, Roosevelt Rd., Taipei, Taiwan 10617
{b91901146, f93942031, b91901119}@ntu.edu.tw, homer@cc.ee.ntu.edu.tw

ABSTRACT

We propose an algorithm to detect depths in a light field. Specifically, given a 4D light field, we find all planes at which objects are located. Although the exact depth of each pixel in the space is left unknown, the partial information obtained is very useful for many applications, such as synthetic aperture photography and all-focused rendering. Our algorithm measures the degree of focus of different planes by calculating the ratio of high frequencies to the low frequencies. To handle different depth distributions, we reformulate the maximum detection problem to a maximum-cover problem that can be solved efficiently by dynamic programming. Compared with auto-focusing and per-pixel depth estimation, our algorithm is much faster yet sufficiently accurate.

Index Terms— Light field, focusing, depth detection, image-based rendering.

1. INTRODUCTION

Before taking a picture with a traditional camera, we need to wait for the camera to determine the best focus point. This automatic focusing (AF) operation has two restrictions. First, there may be many interesting objects at different depths, but it can only choose one to focus on. Second, the objects may be moving during the AF process.

However, AF is no longer necessary with plenoptic camera and integral photography [1], which use optics array to capture 4D light field in one exposure. We can generate images focused at arbitrary depths by transforming and integrating the light field [4]. Some recent developments demonstrate that compact implementation of the plenoptic camera can be achieved by embedding the optics array in conventional cameras [2], [3].

Mathematically, refocusing involves a 2D integration operation in the 4D spatial domain, which tends to be very slow. Nevertheless, this operation is equivalent to sampling a 2D slice in the frequency domain [5], so we can generate various focused photos by efficient slicing and FFT. Furthermore, these focused images can be filtered and combined into an all-focused image [6].

Even with practical devices and fast algorithms, refocusing is still a tedious task because traditional algorithms can only provide one focused depth. If there are multiple objects at different depths in the scene, the users will need to manually select their desired focus.

In this paper we propose a simple but efficient method to solve this problem. Instead of finding the depths of different regions in the scene, we extract the depths of different planes at which objects are located. In other words, for a scene with objects located at (x, y, z)'s, we find the z's of the objects without solving for their x and y positions. Fig. 1 illustrates this idea. As described later, these results are useful for many applications.

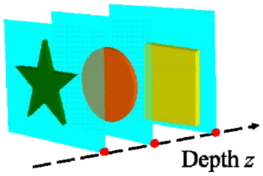


Fig 1. Three objects located at three planes perpendicular to the z axis (optical axis of the camera). Our algorithm finds the depths of these planes.

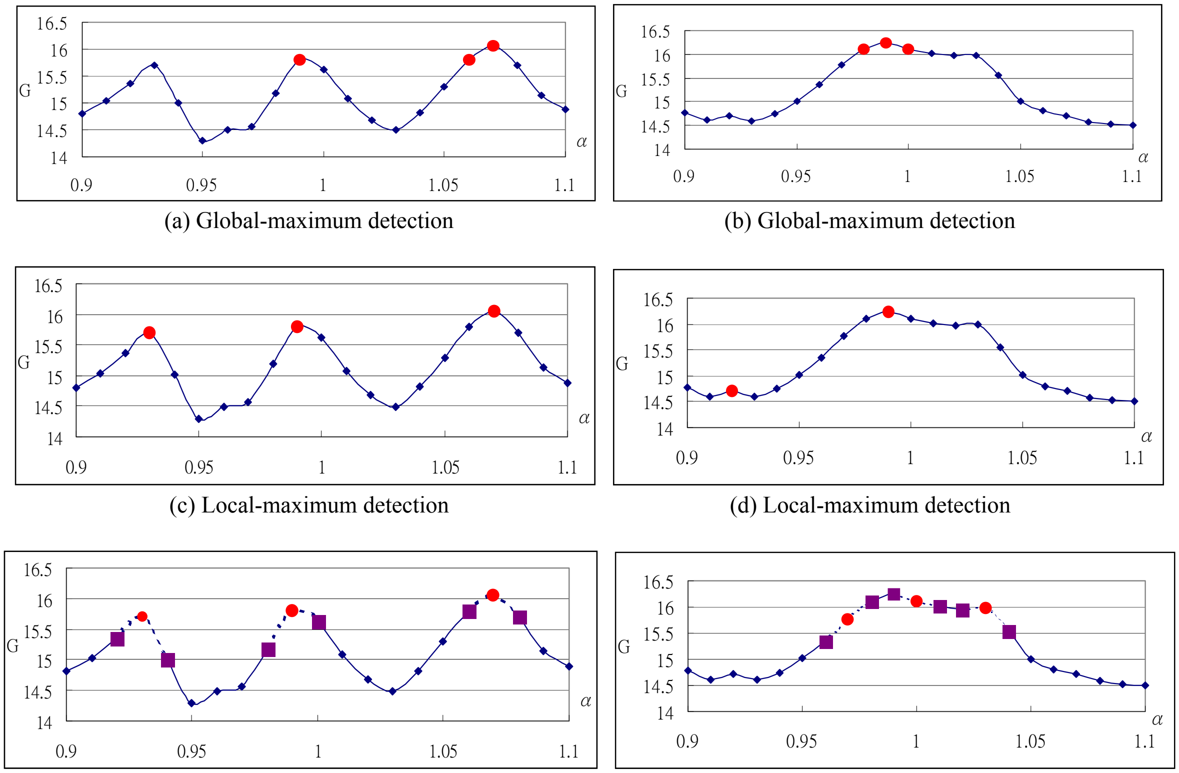
The main concept of our algorithm is based on the observation that the energies of objects at different depths actually lay on different 2D slices in the 4D light field spectrum [8]. Therefore, the ratio of the weighted energy in high frequencies to the energy in low frequencies for these slices provides a cue for detecting the object depths. The computation can be further reduced by considering only two 1D slices for each image plane at different depths. Because the depth distributions vary with content, traditional depth detection methods are not suitable for all scenarios. Here we reformulate the problem and solve it using dynamic programming. Our algorithm entirely operates in frequency domain, so it can integrate with many applications easily. With some modifications, it can operate in the spatial domain as well. The overall complexity of the proposed method is small.

Our proposed method is quite different from traditional AF algorithms in two ways. First, AF algorithms only determine a best focus value for a specific region, but our algorithm detects all depths. Second, AF algorithms are image-based processing and usually use some heuristics to measure the sharpness of the image, while we have a completed 4D light field so we can obtain the depths more precisely.

The organization of this paper is as follows. The proposed algorithm is given in Section 2. Section 3 presents experimental results and two applications: synthetic aperture photography and all-focused light field rendering. The conclusions and future work are drawn in Section 4.

2. PROPOSED ALGORITHM

We say that the *focusness* of a depth is high if there is some object well-focused at the plane of this depth. The basic measurement is presented in Section 2.1. However, objects may spread over a wide range of depths in a real scene, such as ground or forest. In Section



(a) Global-maximum detection

(b) Global-maximum detection

(c) Local-maximum detection

(d) Local-maximum detection

(e) Proposed detection

(f) Proposed detection

Fig. 2. Detected depths of the two datasets using (a) (b) global-maximum, (c) (d) local-maximum, and (e) (f) our proposed algorithm. The circles denote the selected depths, the squares are the neighbors of them, and the dotted segments are the partially focused regions.

2.2 this problem was solved by modifying the depth detection method.

2.1. Measuring the focusness

Assume there is a 4D light field l captured by the plenoptic camera. The depth of the original focal plane is normalized to 1. Note that the depth range is harmonically transformed due to optics. Using *Fast Fourier Transform* (FFT), we can obtain the 4D spectrum L:

$$L\left(f_u, f_v, f_s, f_t\right) = \mathfrak{F}\left\{l\left(u, v, s, t\right)\right\}. \tag{1}$$

From the Fourier slice photography [5], the spectrum $I(\alpha)$ of the synthetic aperture photograph focused at depth α is

$$I\left(f_x, f_y, \alpha\right) = L\left(\left(1-\alpha\right) f_x, \left(1-\alpha\right) f_y, \alpha f_x, \alpha f_y\right). \tag{2}$$

When there is some object located at the plane of depth α, the energy corresponding to the details of this object will mostly fall on the slice of its spectrum [8], so the high frequency component of this spectrum should be larger than those of the other slices which have no object in focus.

To reduce the computation, only the energy along f_x and f_y axes is taken into account. That is, we extract two 1D spectrums from $I(f_x, f_y, \alpha)$:

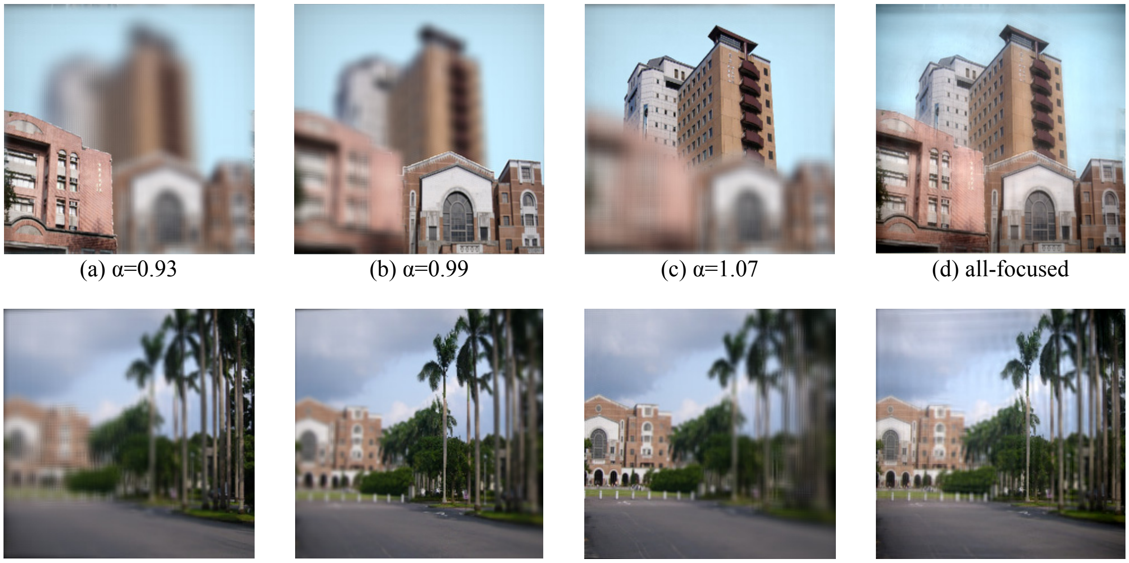
$$\begin{aligned} I_x\left(f_x, \alpha\right) &= L\left(\left(1-\alpha\right) f_x, 0, \alpha f_x, 0\right), \\ I_y\left(f_y, \alpha\right) &= L\left(0, \left(1-\alpha\right) f_y, 0, \alpha f_y\right), \end{aligned} \tag{3}$$

and then calculate the power spectrum $P(f,\alpha)$:

$$P\left(f, \alpha\right) = I_x\left(f, \alpha\right)^2 + I_y\left(f, \alpha\right)^2. \tag{4}$$

The direct summation of $P(f,\alpha)$ over the whole spectrum (total energy) is not a good measurement of focusness for many reasons. First, changing α does not alter the energy of low frequency components. Only the details (high frequencies) are lost when the object is out of focus. Second, noise and aliasing may dominate the energy at high frequencies near the Nyquist rate.

To alleviate these problems, we use a multi-band processing method. The power spectrums $P(f,\alpha)$ is split equally into 8 bands. Denote the energy in these 8 bands by $E_0(\alpha)$, $E_1(\alpha)$,…, $E_7(\alpha)$. The high bands E_1-E_6 are emphasized with proper weighting, and their


(a) α=0.93 (b) α=0.99 (c) α=1.07 (d) all-focused

(e) α=0.97 (f) α=1.00 (g) α=1.03 (h) all-focused

Fig. 3. The synthetic aperture photographs and all-focused photos from synthetic datasets. α denotes the detected depths and all-focused images are generated from the synthetic aperture photographs without human interaction.

summation are normalized by the lowest band E_0. The highest band E_7 is ignored since it contains mostly noise. Denote the measurement of focusness by $G(\alpha)$:

$$G(\alpha)=\frac{1}{\log E_0(\alpha)}\sum_{i=1}^{6} w_i \log E_i(\alpha). \tag{5}$$

We tried many different settings of w_i and found that $w_i = i$ gives the best results.

2.2. The discreteness of the depths

Fig. 2(a) and (b) show the plots of G over α for our two synthetic datasets, which are designed for two extreme cases. It is obvious that $G(\alpha)$ is high when there is some object at depth α. However, simple local or global maximum detection may not result in good selection.

Fig. 2(a) shows the G of dataset 1 with completely discrete depth distribution. The objects in this space are located at three planes with α = 0.93, 0.99, and 1.07. The curve shows three local peaks exactly at these points. On the other hand, Fig. 2(b) shows the G of dataset 2 with completely continuous depth distribution. The objects here spread a range of depths from α = 0.97 to 1.03. The G's in this range are globally higher than those in others, but there is only one peak.

These two extreme cases reveal that naïve depth-detection algorithms based solely on local or global maximums would not work. That is, if we detect the depth by local maximums, we can succeed in the discrete-depth case, but fail to handle the continuous-depth case. On the contrary, the global maximum detection works well in the continuous-depth case, but not in the discrete-depth one.

Note that when there is an object at α, $G(\alpha)$ is larger than its neighbors. In addition, the neighboring G's are also affected by this object. This effect should be taken into account. Instead of finding global or local maximums, we try to solve the following maximum-cover problem:

Given depths α_1, $\alpha_2 ... \alpha_N$, and corresponding G factors G_1, $G_2 ... G_N$, find K indexes D_1, D_2, ..., D_K such that

$$\sum_{i=1}^{K}(\lambda G_{D_i-1}+G_{D_i}+\lambda G_{D_i+1}) \tag{6}$$

is maximized, under the constraint that the selected D_i are separated by at least 3; λ is between 0 and 1.

The constraint is to ensure that the neighbors of every selected depth will not overlap. In our experiments the λ is set to 0.5 and K is set to 3. This problem can be solved efficiently by dynamic programming.

3. EXPERIMENTAL RESULTS

We first perform experiments using the previous synthetic datasets so the exact depths of the objects are known. The resolution for these dataset is 16×16×256×256. For Eq. (3), spectrum is resampled by a Kaiser-Bessel filter with width 2.5. For Eq. (6), the N is set to 21, corresponding to α = 0.90, 0.91, …, 1.10.

Then we perform similar experiments using real dataset captured by programmable aperture camera [9]. The depth distribution of this dataset is neither completely discrete nor completely continuous. Instead, it is composed of an object located at α=0.97 and a region of objects through α=1.01 to α=1.04. The resolution of this dataset is 4×4×256×256.

3.1. Detection of depths

The resulted G curves of the two synthetic datasets are shown in Fig. 2. The global-maximum method fails when the depth distribution is discrete, since the object at 1.07 also pulls up $G(1.06)$, as shown in Fig. 2(a). On the other hand, the local-maximum method fails when the depth distribution is continuous, as shown in Fig. 2(d). There are objects distributed from 0.97 to 1.03, but only $G(0.99)$ is a local peak. For both cases, our proposed algorithm works well. Using the detected depths, we can automatically generate the synthetic aperture photos, as shown in Fig. 3(a-c) and 3(e-g). The results on real dataset are presented in Fig. 4. Our algorithm successfully selects the depths where objects are located at.

We can further generate all-focused images [6] using the synthetic aperture photos, as shown in Fig. 3 (d)(h), and Fig. 4 (d). The little ghosting effect is due to aliasing. Our implementation is more advanced than the previous work in two ways. First, the depths are determined automatically instead of exhaustively. Therefore no user interaction is required. Second, the previous method performs de-convolution in the frequency domain instead of iterations in the spatial domain in [6]. Our method can resolve the magnification in different focused images easily and the computation is reduced significantly. These differences make our system much more efficient.

3.2. Computational cost

We consider the analytic complexity first. For each depth, calculating a single G factor takes O(S), where S is the width (or height) of the image. Calculating N factors takes O(NS). Our selection algorithm takes O(NK), so the total time cost is O(NS+NK) and dominated by O(NS). Compared with the FFT for obtaining L, which takes O($S^4\log S$), and the IFFT for synthetic aperture photo, which takes O($S^2\log S$), the complexity of our algorithm is negligible.

In our experiment, 4D FFT takes 10 seconds, but it can be pre-calculated only once. For some scenario where the 4D FFT is redundant or undesirable, there is a different approach to evaluate Eq. (3). In I_x and I_y, two of the four dimensions are simply DC components. These DC components can be easily extracted by projection. Therefore, we can generate two 2D signals i_{us}, i_{vt} by projecting the 4D signal along v,t and u,s axes respectively, and then perform 2D FFT on i_{us} and i_{vt} to obtain I_x and $I_{y.}$

Our experiments were performed on a PC with Pentium-4 3GHz CPU and 1GB RAM. Generating each refocused image takes 1 second and the all-focused image takes 3 seconds. Our depth detection only takes 0.25 seconds.

4. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed a method to detect all object depths in the light field. The focusness is measured by the weighted summation of energy in different frequency bands, and the detection problem is reformulated into a maximum-cover problem. The experimental results show that our algorithm is fast, accurate, and useful for many applications, such as digital refocusing and all-focused rendering. Compared with other operations in the system, the additional overhead of our method is negligible.

Our algorithm can be applied to many other applications. For example, in multi-camera matting, moving objects must be refocused continuously and automatically [7]. Another one is depth quantization. In [8] a method is proposed to quantize the depth uniformly in disparity. However, many depths may be vacant, so non-uniform quantization will give better results. Also for per-pixel depth estimation, our result can reduce the search space and thus speed up the estimation.

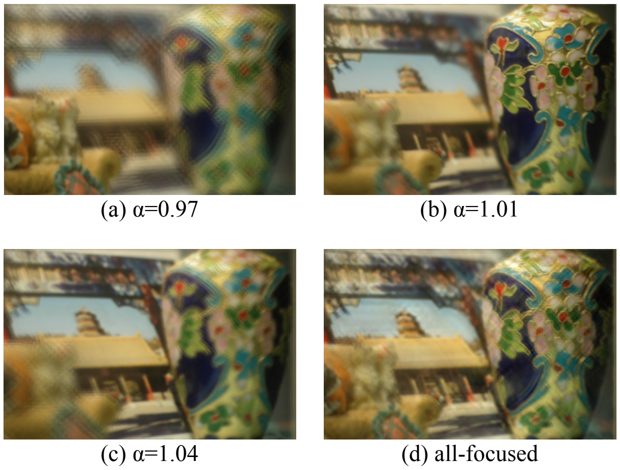


Fig. 4. The synthetic aperture photographs and the all-focused image from the real dataset.

5. ACKNOWLEDGEMENTS

This work was supported in part by grants from the NSC of Taiwan under the contract NSC 94-2752-E-002-006-PAE. Chia-Kai Liang was supported by the MiTac Technology Scholarship.

6. REFERENCES

[1] E. H. Adelson and J. Y.A. Wang, “Single lens stereo with plenoptic camera,” in *IEEE Trans. PAMI*, vol. 14, no. 2, pp. 99-106, Feb. 1992.

[2] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, “Light field photography with a hand-held plenoptic camera,” *Stanford University CSTR 2005-02*, 2005.

[3] T. Georgeiv, K. C. Zheng, B. Curless, D. Salesin, S. K. Nayar, and C. Intwala, “Spatio-angular resolution tradeoff in integral photography,” in *Proc. EGSR*, 2006.

[4] A. Isaksen, L. McMillan, and S. J. Gortler, “Dynamically reparameterized light fields,” in *Proc. SIGGRAPH'00*, pp. 297-306, 2000.

[5] R. Ng, “Fourier slice photography,” in *ACM Trans. Graph (Proc. SIGGRAPH’05)*, vol. 24, no. 3, pp. 735-744, Jul. 2005.

[6] A. Kubota, K. Takahashi, K. Aizawa, and T. Chen, “All-focused light field rendering,” in *Proc. Eurographics Symposium on Rendering*, June 21-23, 2004.

[7] N. Joshi, W. Matusik, and S. Avidan, “Natural video matting using camera arrays,” in *ACM Trans. Graph (Proc. SIGGRAPH’06)*, vol. 25, no. 3, pp. 779-786, Jul. 2006.

[8] J.-X. Chai, S.-C. Chan, H.-Y. Shum, and X. Tong, “Plenoptic sampling,” in *Proc. SIGGRAPH'00*, pp. 307-318, 2000.

[9] C.-K. Liang, G. Liu, and H. H. Chen, “Light field acquisition using programmable aperture camera,” to be submitted.