Multilayer Image Disparity Estimation and Blending for Light Field Cameras

Shih-Chang Chuang, Jian-Jiun Ding *, and Po-Jen Chen
Graduate Institute of Communication Engineering, National Taiwan University
E-mail: r02942121@ntu.edu.tw, jjding@ntu.edu.tw, r03942106@ntu.edu.tw
Tel: +886-2-33669652

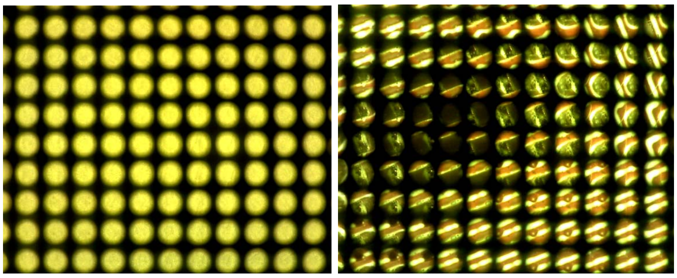
***Abstract*—Due to the development of hand-held Plenoptic cameras, the light filed camera has become more and more popular in recent years. The light field camera can capture the angular and depth information and is able to construct a stereoscopic image. To obtain a high quality reconstructed image, the disparity should be estimated accurately. However, for the sub-images taking from the light field camera, especially for those of the microscope case, it usually happens that there is no feature point within a patch and the disparity is hard to be determined accurately. In this paper, to improve the efficiency and the accuracy of disparity estimation for light field cameras, we propose a multilayer matching architecture. Moreover, a gradient map is applied to assign the weight for disparity matching. In addition, we apply the segmentation method to the rendered final image and refine the disparity estimation result. Simulations show that a more accurate depth estimation result can be achieved when using the proposed architecture.**

I. Introduction

The light field camera becomes more and more popular in recent years. It is useful for constructing a 3D image and can estimate the depths of objects. One can use a light field camera to obtain a set of micro images, as in Fig. 1. After determining the matched patches of two or more micro images, as in Fig. 2, the disparity and hence the depth can be calculated from the relative locations among the matched patches.

In [1], a simple disparity determination and depth estimation method based on correlation operations was proposed. In [2], they assumed that the objects are Lambertian ones and use it to construct the depth map. However, it may not be not suitable for the focused light field camera. In [3, 4], another methods based on epi-polar plane images (EPIs) were proposed. The EPIs can be viewed as the 2D slices of a lumi-graph captured by a camera. To generate EPIs, one can extract the same row or column in sequential images and then arrange them in a sequential order. The angles of the line structure in EPIs is proportional to the depth of the object. However, to obtain an accurate depth map, the samples should be dense enough. Moreover a method based on adaptive normalized cross-correlation was proposed in [5].

In this paper, to improve the accuracy of depth estimation, a multilayer architecture is proposed. Note that, in a micro image, it usually happens that there is no (or very few) feature point within a patch, which limits the performance of disparity estimation.



(a) (b)

Fig. 1 (a) The white image. (b) A micro image set taken from the light filed camera. These images were taken by the Industrial Technology Research Institute, Taiwan, with a 2.0 Plenoptic light field camera.

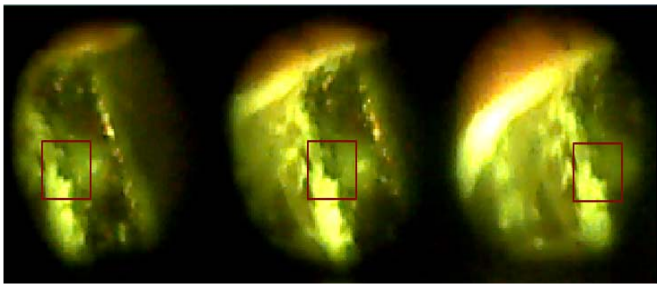


Fig. 2 The matched patches of micro images.

Therefore, we propose a multilayer architecture to solve the problem. First, large-scale patches, which are the combinations of several micro images, are adopted to roughly estimate the disparity. Then, in the 2^{nd} to the 4^{th} layers, the patches become smaller and smaller. We then use these smaller patches together with the disparity estimated from large-scale patches to determine the disparity accurately.

Moreover, we think that the pixel with larger gradients should play more important roles for disparity estimation. Therefore, we use the gradient to determine the weight and apply the weighted correlation operation to estimate the disparity and the depth.

Furthermore, we apply the Gaussian filter and regression analysis to precisely determine the shift of micro images. We also apply post segmentation method and calculate the average disparity for each object. Simulations show that, with the proposed algorithm, the disparity can be estimated precisely and a more accurate depth map can be achieved.

II. Proposed Method

The flowchart of the proposed algorithm is plotted in Fig. 3.

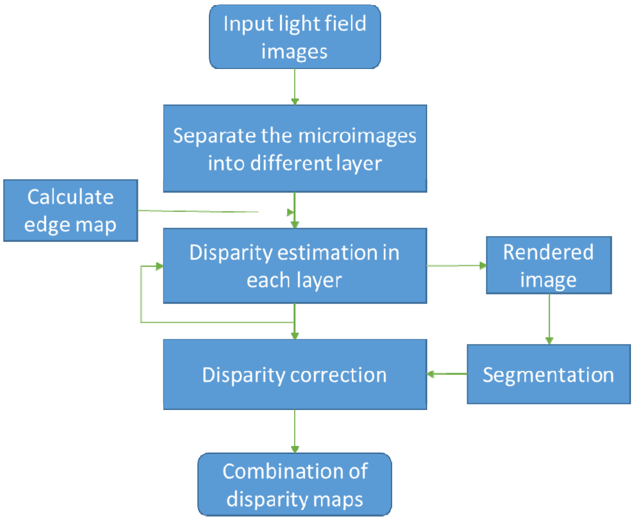


Fig. 3 The architecture of the proposed algorithm.

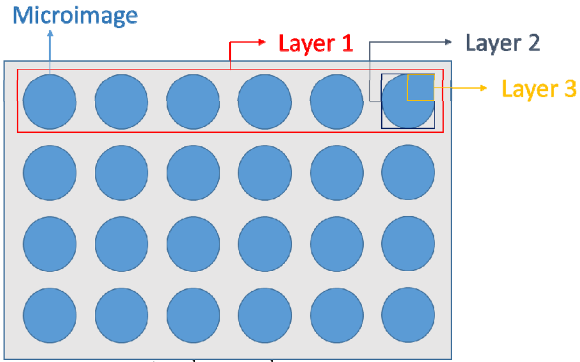


Fig. 4 The 1st, 2nd, and 3rd layers for disparity estimation.

First, we calculate the gradients of micro images and construct an edge map. Then, a compensation operation is applied to reduce the attenuation in surrounding pixels. Furthermore, we apply a multilayer architecture to estimate the disparity accurately. Its detail is described in the following subsection.

A. Multilayer Disparity Estimation

We apply a four-layer architecture for disparity estimation. The patch in the first layer consists of the micro images in the same row (or column), as in Fig. 4. The patch in the 2nd layer is a single sub image. The area of the patch in the 3rd layer is 1/25 of that of the sub image. In the 4th layer, each patch consists of a smaller number of pixels. The disparity of the previous layer can be used to determine the disparity of the next layer. If the estimated disparity of the previous layer of d_0, then in the next layer we can estimate that the disparity is in the range of $[d_0 - T, d_0 + T]$.

We adopt the multilayer discrete parity estimation method because neither the large patch nor the small patch can obtain good disparity estimation result. If the patch is too large, the resolution of the depth map will be limited. If the patch is too small, it may happen that the feature within a patch is too less and the patch with similar colors will be misidentified as a matched patch, as the example in Fig. 5.

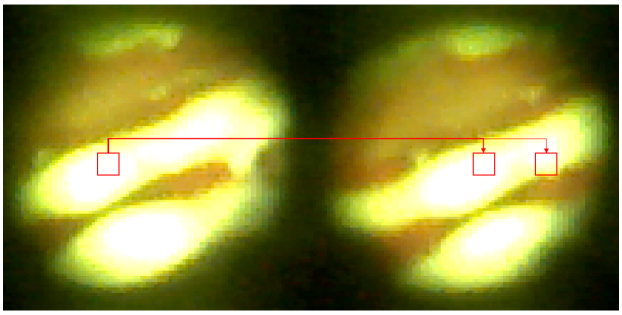


Fig. 5 The red block of the left micro image may match both the two red blocks of the right micro image because their colors are similar.

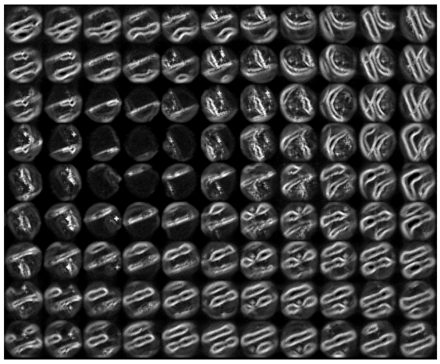


Fig. 6 The gradient map of each micro image in Fig. 1(b).

Therefore, to obtain an accurate disparity estimation result, both large patches and small patches should be applied.

B. Disparity Estimation in Each Layer

First, we apply the white image to compensate the intensity of the borders of micro images. Then, instead of conventional edge detection filters, we apply (1) to determine the gradient It can well detect both the step-type and the ridge-type edges:

$$\begin{aligned} x_g[m,n] &= |x[m,n]*f[n]| + |x[m,n]*f[-n]| \\ &\quad + |x[m,n]*f[m]| + |x[m,n]*f[-m]| \end{aligned} \quad (1)$$

where $f[n] = \left[1 - 2\pi\left(\frac{n}{4}\right)^2\right] * e^{\left[-\pi\left(\frac{n}{4}\right)^2\right]}$ *for* $0 \le n \le 7$.

We show t $x_g[m, n]$ corresponding to Fig. 1(a) in Fig. 6. Then, we assign the weight function as

$$W_\tau[m,n] = w_{1,a}[m, n+\tau] w_{1,b}[m, n+\tau],$$

$$w_1[m,n] = 1 + x_g[m,n]/300$$

and $w_{k,a}[m, n]$ (k = 1 or 2) are the weight functions $w_1[m, n]$ correspond to the two patches. Then, we can calculate the weighted difference sum as

$$c_h[\tau] = \sum_{RGB}\sum_{n}\sum_{m} \left|A_b[m, n+\tau] - A_a[m,n]\right| W_\tau[m,n] \quad (2)$$

where $A_a[m, n]$ and $A_b[m, n]$ are the two patches. Suppose that d_0 is the disparity measured in the previous layer. Then the disparity between $A_a[m, n]$ and $A_b[m, n]$ can be determined from:

$$\arg\min_{\tau} c_h[\tau](1 + (\tau - d_0)^2), \quad (3)$$

$$d_0 - T \le \tau \le d_0 + T.$$

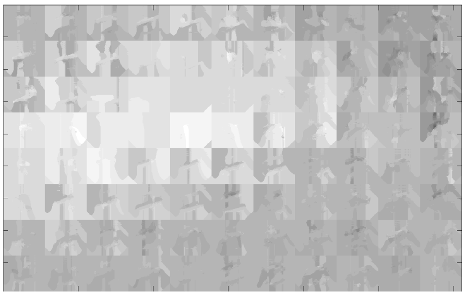


Fig. 7 The disparity map obtained by the process in Section II-B.

In Fig. 7, we show the disparity map for Fig. 1(c) using the proposed algorithm. The bright part means larger disparity and the dark part means smaller disparity. We combine the window based stereo matching algorithm, segmentation method, and use the weighted function determined from white image and the quad-tree architecture to improve the accuracy. The result can be further refined by the disparity correction process and post-segmentation.

C. Disparity Correction

In order to find out the shift of each micro lens, we use the stereo matching result which we computed in the previous subsection. Because we have the disparity map of each microimages, and the microimages is overlapped with adjacent microimage, we can use these disparity map to correct the error of stereo matching caused from occlusion or the inaccuracy of disparity assignment. The process is described as follows.

Let $I_{i,j}(x, y)$ and $D_{i,j}(x, y)$ be the $(i, j)^{th}$ microimage and the disparity map obtained in Section II-B, respectively, and x, y are the coordinates in the microimage. We first compare the disparity map $D_{i,j}(x, y)$ with the disparity maps of adjacent micro images $D_{i-1,j}(x, y)$, $D_{i+1,j}(x, y)$, $D_{i,j-1}(x, y)$, and $D_{i,j+1}(x, y)$. Then, we use two checking processes to avoid the effects of occlusion or wrong disparity on the final disparity map. We calculate

$$M_{m,n} = \begin{cases} 1\ ,\ if\ |I_{i,j}(x,y) - I_{i+m,j+n}\left(x - D_{i,j}(x,y)*m, y - D_{i,j}(x,y)*n\right)| < k\ , \\ 0,\ \ if\ |I_{i,j}(x,y) - I_{i+m,j+n}\left(x - D_{i,j}(x,y)*m, y - D_{i,j}(x,y)*n\right)| > k\ , \end{cases} \tag{4}$$

$$N_{m,n} = \begin{cases} 1, |D_{i,j}(x,y) - D_{i+m,j+n}\left(x - D_{i,j}(x,y)*m, y - D_{i,j}(x,y)*n\right)| < l \\ 0, |D_{i,j}(x,y) - D_{i+m,j+n}\left(x - D_{i,j}(x,y)*m, y - D_{i,j}(x,y)*n\right)| > l \end{cases} \tag{5}$$

where $(m, n) \in \{(-1, 0), (1, 0), (0, 1), (0, -1)\}$. In other words, if the colors of the corresponding points in the adjacent micro images (i.e., $I_{i+m,j+n}\left(x - D_{i,j}(x,y)*m, y - D_{i,j}(x,y)*n\right)$) are not similar to the color of the current pixels, then the corresponding points will not be considered as valid matching points. Then, we correct the disparity map by the valid correspond points of adjacent micro images:

$$D_{i,j}(x,y) = \frac{\sum_{m,n} D_{i,j}(x,y) + \sum_{m,n} D_{i+m,j+n}(x,y)*\left(M_{m,n}*N_{m,n}\right)}{1+\sum_{m,n}\left(M_{m,n}*N_{m,n}\right)} \tag{6}$$

D. Rendered Image

After all the disparity maps are estimated, we combine the micro-images into one image. Suppose that the vertical and horizontal disparities of the p^{th} micro-image $A_p[m, n]$ are (τ_p, η_p) (respected to the leftmost and upmost micro-image) and its weight function of calculated from the white input is $w_{2,p}[m, n]$. Then, we separate the sub images into three levels (high frequency, middle frequency and low frequency levels) and the rendered image for each level is determined by

$$B[m,n] = \sum_p A_p[m-\tau_p, n-\eta_p] \frac{W_{2,p}^{\alpha}[m-\tau_p, n-\eta_p]}{\sum_p W_{2,p}^{\alpha}[m-\tau_p, n-\eta_p]} \tag{7}$$

After combining the results for each level and each segment, we obtained the rendered image.

E. Segmentation

The disparities of the pixels within a region should be a constant or a continuous function. Therefore, it is proper to perform image segmentation before obtaining the final disparity map. After image blending, we apply image segmentation to separate the rendered image into several parts. We suggest that the superpixel-based segmentation method [14], which can make the boundaries of each region highly match the edges of objects, can be used for segmentation.

F. Combination of Disparity Maps

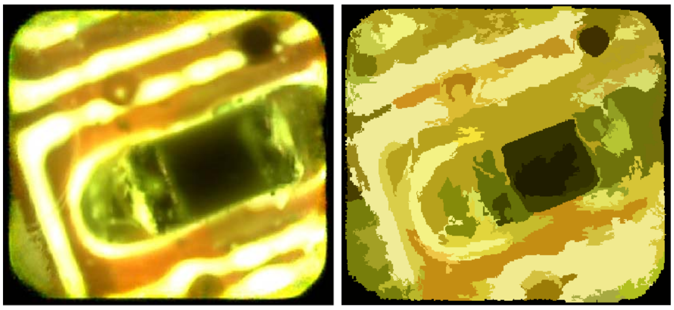
We have calculated the disparity value of each pixel in Section II-C. Moreover, which region each pixel belongs to has been determined in Section II-E. Then, we can use these results to calculate the average disparity value for each region. After determining the disparity, the depth can be determined by the parameters of cameras.

III. Simulations

In this section, simulation results are shown. For the light field image in Fig. 1(b), which was a picture of a transistor acquired from a microscope light field camera, the disparity map obtained from the process in Section II-B, the rendering result, and the segmentation result are shown in Figs. 7, 8(a), and 8(b), respectively. Moreover, in Fig. 9, we show the final depth map for each region. We apply the gray level to show the depth and higher intensity means that the distance is smaller.

The results in Fig. 9 shows that the green color part in Fig. 8 has a smaller distance and the surrounding part has a larger distance. It matches the fact that the cental part of the transistor is near to the lenses of the microscope light field camera. By contrast, when using the methods of discrete correlation [1] + segmentation, and adaptive normalized cross-correlation [5], the depth maps are shown in Figs. 10 and 11, respectively. The fact that the central part of the transistor has a smaller distance cannot be seen from Figs. 10 and 11.

IV. Conclusion



(a) (b)

Fig. 8 (a) The rendered image, (b) the segmentation result.

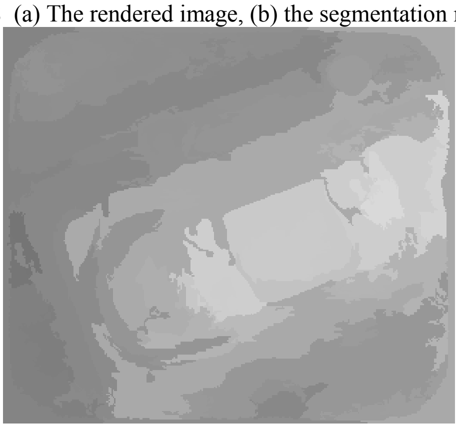


Fig. 9 Depth map of the microscope light field image in Fig. 1(b) using th PROPOSED algorithm. It matches the fact that the central part of the convex transistor is near to the camera and the surrounding part is far from the camera.

In this paper, we propose a disparity and depth estimation algorithm for light field cameras. It adopts the techniques of multilayer disparity estimation, the gradient-based weight function, and post segmentation. Compared with the other light field depth reconstruction methods, the proposed method can achieve a more accurate result. For example, for the convex transistor in Fig. 8(a), although the depths for different parts have very small differences, when using the proposed algorithm, one can validly identify that the central part has a smaller distance to the camera than the surrounding part.

Fig. 10 Depth map for Fig. 1(b) when using the correlation algorithm [1] and post segmentation

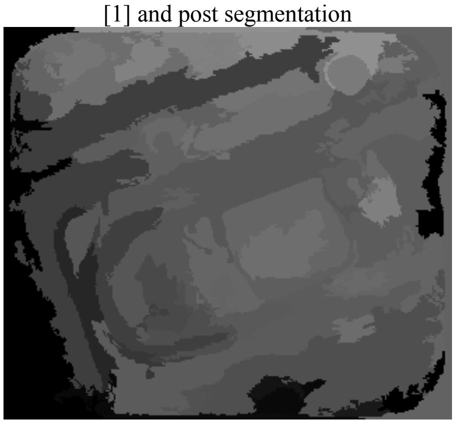


Fig. 11 Depth map for Fig. 1(b) when using adaptive normalized cross-correlation [5] and post segmentation

REFERENCES

[1] T. Georgive and A. Lumsdainea, “Focused plenoptic camera rendering”, *Journal of Electronic Imaging*, vol. 19, issue 2, pp. 021106-021106-11, 2010.

[2] T. E. Bishop and P. Favaro “Plenoptic depth estimation from multiple aliased views,” *ICCV Workshops*, pp. 1622-1629, 2009

[3] R. C. Bolles, H. H. Baker, and D. H. Marimont. "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, issue 1, pp. 7-55, 1987.

[4] S. Wanner and B. Goldluecke, "Globally consistent depth labeling of 4D light fields," *IEEE Conference on Computer Vision and Pattern Recognition,* pp. 41-48, June 2012.

[5] Y. S. Heo, M. L. Kyoung, and S. U. Lee. "Robust stereo matching using adaptive normalized cross-correlation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 33, issue 4, pp. 807-822, 2011.

[6] E. H. Adelson and J. R. Bergen, *The Plenoptic Function and the Elements of Early Vision*, Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology, 1991.

[7] M. Levoy and P. Hanrahan, "Light field rendering," *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, ACM*, pp. 31-42, 1996.

[8] B. Wilburn, N. Joshi, V. Vaish, E. V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, issue 3, pp. 765-776, 2005.

[9] M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report*, vol. 2, pp. 1-11, 2005.

[10] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.*, vol. 26, issue 3, article 69, 2007.

[11] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, issue 3, article 70, 2007.

[12] C. K. Liang, T. H. Lin, B. Y. Wong, C. Liu, and H. H. Chen, "Programmable aperture photography: multiplexed light field acquisition," *ACM Trans. Graph.*, vol. 27, article 55, 2008.

[13] A. Lumsdaine and T. Georgiev. "Full resolution lightfield rendering," *Indiana University and Adobe Systems, Tech. Rep.*, 2008.

[14] Z. Li, X. M. Wu, and S. F. Chang, “Segmentation using superpixels: A bipartite graph partitioning approach,” in *CVPR*, pp. 789-796, 2012.