Accurate Depth Map Estimation from a Lenslet Light Field Camera

Hae-Gon Jeon, Jaesik Park, Gyeongmin Choe, Jinsun Park, Yunsu Bok, Yu-Wing Tai and In So Kweon
Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea

The problem of estimating an accurate depth map from a lenslet light field camera, e.g. Lytro™[1] and Raytrix™[2], is investigated. Because the baseline between sub-aperture images from a lenslet light field camera is very narrow, directly applying the existing stereo matching algorithms such as [3] cannot produce satisfying results. In this paper, an algorithm for stereo matching between sub-aperture images with an extremely narrow baseline is presented. Central to the proposed algorithm is the use of the phase shift theorem in the Fourier domain to estimate the sub-pixel shifts of sub-aperture images. This enables the estimation of the stereo correspondences at sub-pixel accuracy, even with a very narrow baseline. In addition, a method of correcting distortions of lenslet light field cameras is also presented.

Distortion Estimation and Correction During the capture of a light field image of a planar object, spatially variant epipolar plane image (EPI) slopes (i.e. non-uniform depths) are observed that result from the distortions. In addition, the degree of distortion also varies for each sub-aperture image. To solve this problem, an energy minimization problem is formulated under a constant depth assumption as follows:

$$\hat{G} = \underset{G}{\operatorname{argmin}} \sum_{\mathbf{x}} |\theta(I(\mathbf{x})) - \theta_o - G(\mathbf{x})| \tag{1}$$

where $|\cdot|$ denotes the absolute operator. θ_o, $\theta(\cdot)$, and $G(\cdot)$ denote the slope without distortion, the slope of EPI, and the amount of distortion at point $\mathbf{x}$, respectively. An image of a planar checkerboard is captured and compared with the observed EPI slopes with θ_o. Points with strong gradients in the EPI are selected and the difference $(\theta(\cdot) - \theta_o)$ is calculated in Eq. (1). Then, the entire field curvature G is fitted to a second order polynomial surface model.

Depth Map Estimation Given the distortion-corrected sub-aperture images, the goal is to estimate accurate dense depth maps. The proposed depth map estimation algorithm is developed using a cost-volume-based stereo [3]. In order to manage the narrow baseline between the sub-aperture images, the pipeline is tailored with three significant differences described in the next.

1) Phase Shift based Sub-pixel Displacement A key contribution of the proposed depth estimation algorithm is matching the narrow baseline sub-aperture images using sub-pixel displacements. According to the phase shift theorem, if an image I is shifted by $\Delta\mathbf{x} \in \mathbb{R}^2$, the corresponding phase shift in the 2D Fourier transform is:

$$\mathcal{F}\{I(\mathbf{x}+\Delta\mathbf{x})\} = \mathcal{F}\{I(\mathbf{x})\}\exp^{2\pi i \Delta\mathbf{x}}, \tag{2}$$

where $\mathcal{F}\{\cdot\}$ denotes the discrete 2D Fourier transform. In Eq. (2), multiplying the exponential term in the frequency domain is the same as convolving a Dirichlet kernel (or periodic sinc) in the spatial domain. According to the Nyquist-Shannon sampling theorem [4], a continuous band-limited signal can be perfectly reconstructed through convolving it with a sinc function. If the centroid of the sinc function is deviated from the origin, precisely shifted signals can be obtained. In the same manner, Eq. (2) generates a precisely shifted image in the spatial domain if the sub-aperture image is band-limited. Therefore, the sub-pixel shifted image $I'(\mathbf{x})$ is obtained using:

$$I'(\mathbf{x}) = I(\mathbf{x}+\Delta\mathbf{x}) = \mathcal{F}^{-1}\{\mathcal{F}\{I(\mathbf{x})\}\exp^{2\pi i \Delta\mathbf{x}}\}. \tag{3}$$

2) Building the Cost Volume In order to match sub-aperture images, two complementary costs were used: the sum of absolute differences (SAD) and the sum of gradient differences (GRAD). The cost volume C is defined as a function of $\mathbf{x}$ and cost label l:

$$C(\mathbf{x},l) = \alpha C_A(\mathbf{x},l) + (1-\alpha) C_G(\mathbf{x},l), \tag{4}$$

Lytro software [1] Ours

Figure 1: Synthesized views of the two depth maps acquired from Lytro software [1] and our approach.

where $\alpha \in [0,1]$ adjusts the relative importance between the SAD cost C_A and GRAD cost C_G which are defined as

$$\begin{aligned} C_A(\mathbf{x},l) &= \sum_{\mathbf{s}\in V} \sum_{\mathbf{x}\in R_\mathbf{x}} \min(|I(\mathbf{s}_c,\mathbf{x}) - I(\mathbf{s},\mathbf{x}+\Delta\mathbf{x}(\mathbf{s},l))|, \tau_1), \\ C_G(\mathbf{x},l) &= \sum_{\mathbf{s}\in V} \sum_{\mathbf{x}\in R_\mathbf{x}} \beta(\mathbf{s}) \min\left(\mathrm{Diff}_x(\mathbf{s}_c,\mathbf{s},\mathbf{x},l), \tau_2\right) \\ &\quad + \left(1-\beta(\mathbf{s})\right) \min\left(\mathrm{Diff}_y(\mathbf{s}_c,\mathbf{s},\mathbf{x},l), \tau_2\right) \end{aligned} \tag{5}$$

where $R_\mathbf{x}$ is a small rectangular region centered at $\mathbf{x}$; τ_1 and τ_2 are truncation values of robust functions, V contains the st coordinate pixels $\mathbf{s}$, except for the center view $\mathbf{s}_c$ and $\mathrm{Diff}_x(\mathbf{s}_c,\mathbf{s},\mathbf{x},l) = |I_x(\mathbf{s}_c,\mathbf{x}) - I_x(\mathbf{s},\mathbf{x}+\Delta\mathbf{x}(\mathbf{s},l))|$ denotes the differences between the x-directional gradient of the sub-aperture images. $\beta(\mathbf{s})$ controls the relative importance of the two directional gradient differences based on the relative st coordinates and is defined as $\beta(\mathbf{s}) = \frac{|s-s_c|}{|s-s_c|+|t-t_c|}$. Equation (2) is used for precise sub-pixel shifting of the images. Equation (5) builds a matching cost through comparing the center sub-aperture image $I(\mathbf{s}_c,\mathbf{x})$ with the other sub-aperture images $I(\mathbf{s},\mathbf{x})$ to generate a disparity map from a canonical viewpoint. The 2D shift vector $\Delta\mathbf{x}$ in Eq. (5) is defined as $\Delta\mathbf{x}(\mathbf{s},l) = lk(\mathbf{s}-\mathbf{s}_c)$ where k is the unit of the label in pixels. $\Delta\mathbf{x}$ linearly increases as the angular deviations from the center viewpoint increase.

3) Disparity Optimization and Enhancement In order to reduce the effects of image noise, the weighted median filter was adopted to remove the noise in the cost volume, followed by using the multi-label optimization to propagate reliable disparity labels to the weak texture regions. In the multi-label optimization, confident matching correspondences between the center view and other views are used as additional constraints, which assist in preventing oversmoothing at the edges and texture regions. Finally, the estimated depth map is iteratively refined using quadratic polynomial interpolation to enhance the estimated depth map with sub-label precision.

[1] The lytro camera. http://www.lytro.com/.

[2] Raytrix. 3d light field camera technology. http://www.raytrix.de/.

[3] Christoph Rhemann, Asmaa Hosni, Michael Bleyer, Carsten Rother, and Margrit Gelautz. Fast cost-volume filtering for visual correspondence and beyond. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[4] Claude E. Shannon. Communication in the presence of noise. *Proceeding of the IEEE*, 86(2):447–457, 1998.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.