The 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI 2015)
October 28 ∼ 30, 2015 / KINTEX, Goyang city, Korea

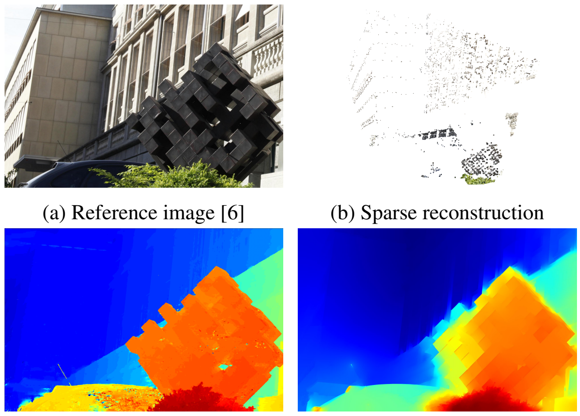
Depth Estimation from Light Field Cameras

Sunghoon Im[1], Hae-Gon Jeon[2], Hyowon Ha[3] and In So Kweon[4]

Korea Advanced Institute of Science and Technology, Republic of Korea
(E-mail: shim, hgjeon, hwha@rcv.kaist.ac.kr, iskweon77@kaist.ac.kr)

Abstract - We present a 3D reconstruction method to estimate depth maps using light-field camera. The estimated depth is widely used for photographic editing. The proposed algorithm, as a feature-based method, provides high quality depth map. First, the distinctive features are extracted from reference image, and the correspondences are tracked. Second, the depth values for each points are calculated by small angle approximated bundle adjustment. Finally, the whole depth map is estimated by simple depth propagation method. To show the effectiveness of our algorithm, we compare our results with the state-of-the-art algorithm.

Keywords - Depth estimation, Light-field, Propagation.



(a) Reference image [6] (b) Sparse reconstruction

(c) Depth map from [6] (d) Our depth map

Fig. 1 System overview (Statue : 2622x1718).

1. Introduction

Recently, the depth can be used for many user-friendly applications, such as refocusing and parallax. Many companies and academics have concentrated much effort on computing accurate depth map. Among various approaches, such as light-field, stereo camera, SfM(Structure from Motion), light-field imaging become powerful solution. Lytro and Pelican launch their depth estimation camera, and provide interesting applications. Because of users' needs for more realistic application, many researchers [1, 2] have studied on light-field imaging. Bok *et al.* [1] study on the method to accurately calibrate light-field camera, and Jeon *et al.* [2] obtain depth from micro-lens-based light-field camera. In this paper, we propose high-quality depth estimation method from high-resolution light-field images. There are mainly three steps: feature extraction, sparse points reconstruction, and depth propagation. We follow 3D reconstruction method from accidental motion [3] and show that the algorithm can also estimate high quality depth from light-field images. As shown in Fig. 1, our depth map outperforms depth map from the state-of-the-art algorithm.

2. Proposed Method

2.1 Feature extraction

Feature extraction is the most essential parts to obtain high quality depth map from light-field images. First, the corner features are extracted from the reference view, and they are tracked by KLT tracker [4]. Because KLT tracker provides subpixel level coordinates, it is powerful solution, but it also tracks features on moving objects. These features make serious artifacts and it should be removed in advance. Therefore, we use RANSAC-based fundamental matrix between reference and the other frames to remove the features on moving object. When the feature is assigned as outlier, we do not consider the feature for next sparse 3D reconstruction step.

2.2 Sparse points reconstruction

Im *et al.* propose the bundle adjustment specialized in small motion. It is defined as:

$$F = \sum_{i=1}^{N_C} \sum_{j=1}^{N_P} ||p_{ij} - \pi(K(R_i P_j + T_i))||^2, \qquad (1)$$

where N_C and N_P are the number of cameras and feature points. R and T are camera rotation matrix and translation vector respectively. p and P are 2D image and 3D world coordinates, and K and π are camera intrinsic matrix and projection function respectively. Because this bundle adjustment is convex in the specific range [5], we initialize depth as random point depth. Additionally, we initialize rotations and translations as zero rotation and translation.

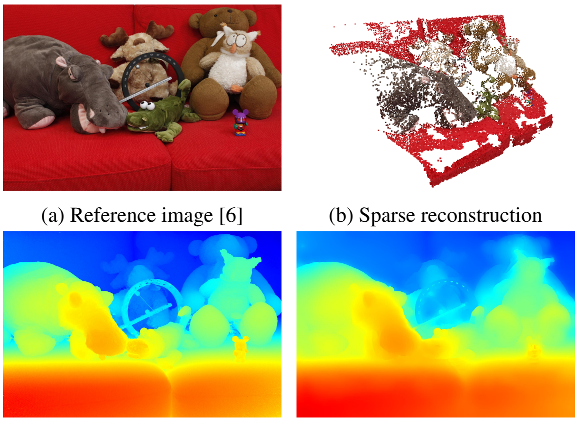
Im *et al.* argue that the bundle adjustment is only for small motion, but it is only specialized in small angle, not translation. Therefore, the bundle adjustment can be adjusted to light-field images which have wide translation [6]. As shown in Fig. 1-(b), the depth of sparse points are estimated, and it will be initial cue for whole depth estimation.

2.3 Depth propagation

With the sparse points from Sec 2.2, we obtain dense depth map by using color affinity [7]. The cost function is designed by data term E_d and smoothness term E_c:

$$E(D) = E_d(D) + \lambda E_c(D), \qquad (2)$$

where, D is the depth map that we want to obtain, and λ is regularization parameter. The meaning of data term is



(a) Reference image [6] (b) Sparse reconstruction

(c) Depth map from [6] (d) Our depth map

Fig. 2 Result comparison (Couch : 4020x2679).

that the target depth at pixel j should be similar with the depth of sparse 3D points.

$$E_d(\mathbf{D}) = \sum_{\mathbf{j}} \left(\mathbf{D_j} - \mathbf{Z_j} \right)^{\mathbf{2}}, \tag{3}$$

where $\mathbf{Z}$ is known depth from Sec 2.2.

The color smoothness term is defined as:

$$E_c(\mathbf{D}) = \sum_{\mathbf{p}} \sum_{\mathbf{q} \in \mathbf{N_8}} \left(\mathbf{D_p} - \frac{\mathbf{w^c_{pq}}}{\sum_{\mathbf{q}} \mathbf{w^c_{pq}}} \mathbf{D_q} \right)^{\mathbf{2}}, \tag{4}$$

where q is the 8-neighboring pixel of q, and p is the pixel that we want to estimate depth. The weight w is defined as:

$$w^c_{pq} = \exp\left(\sum_{\mathbf{I} \in RGB} -\frac{|\mathbf{I_p} - \mathbf{I_q}|}{2\sigma_p^2} \right), \tag{5}$$

$$where\ \sigma_p^2 = \sum_{q \in N_8} (\mathbf{I_p^2} - \mathbf{I_q^2}), \tag{6}$$

where, I is the intensity of images, and σ is the variance of the neighboring intensity. As the weight is color affinity, it has high affinity when the color of neighboring pixel is similar with the color of center point. The cost function can be linearly solved with the equation:

$$\nabla E(D) = 0. \tag{7}$$

3. Experimental Result

We use 10 light-field images and intrinsic matrix provided from [6], and they also provide depth map from their algorithm. We set the regularization parameter λ as 3. Additionally, we run our algorithm on Intel i7 3.40GHz CPU and 16GB RAM, and it takes about five minutes on MATLAB. As shown in Fig 2, the depth map from our algorithm outperforms depth from the state-of-the-art [6]. Especially, our depth map is less noisy, and preserve more details.

4. Conclusion and Discussion

We show that the bundle adjustment for small motion is also applied to wide-translation light-field setup. We

Fig. 3 Application : Refocusing.

obtain high quality depth map from light-field images and it can be widely used for various application shown in Fig 3. In the future, we may adapt our algorithm to micro-lens-based light-field camera. Because it has extremely narrow-baseline, the depth uncertainty is much high, and it may be hard problem.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2010-0028680).

References

[1] Yunsu Bok, Hae-Gon Jeon, and In So Kweon, “Geometric calibration of micro-lens-based light-field cameras using line features,” in *Proc. of European Conf. on Computer Vision (ECCV)*. 2014.

[2] Hae-Gon Jeon, Jaesik Park, Gyeongmin Choe, Jinsun Park, Yunsu Bok, Yu-Wing Tai, and In So Kweon, “Accurate depth map estimation from a lenslet light field camera,” in *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2015.

[3] Sunghoon Im, Gyeongmin Choe, Hae-Gon Jeon, and In So Kweon, “Depth from accidental motion using geometry prior,” in *IEEE International Conference on Image Processing (ICIP)*. 2015.

[4] Carlo Tomasi and Takeo Kanade, *Detection and tracking of point features*, School of Computer Science, Carnegie Mellon Univ. Pittsburgh, 1991.

[5] Fisher Yu and David Gallup, “3d reconstruction from accidental motion,” in *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2014.

[6] Changil Kim, Henning Zimmer, Yael Pritch, Alexander Sorkine-Hornung, and Markus H Gross, “Scene reconstruction from high spatio-angular resolution light fields.,” *ACM Trans. Graph.*, vol. 32, no. 4, pp. 73, 2013.

[7] Jaesik Park, Hyeongwoo Kim, Yu-Wing Tai, Michael S Brown, and Inso Kweon, “High quality depth map upsampling for 3d-tof cameras,” in *Proc. of Int’l Conf. on Computer Vision (ICCV)*, 2011.