A SIMPLE AND ROBUST SUPER RESOLUTION METHOD FOR LIGHT FIELD IMAGES

Yunlong Wang[1,2], *Guangqi Hou*[2], *Zhenan Sun*[2], *Zilei Wang*[1], *Tieniu Tan*[2]

[1]University of Science and Technology of China
[2]Center for Research on Intelligent Perception and Computing,
National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences

ABSTRACT

Light field cameras generate low-resolution images due to the tradeoff between spatial and angular resolution. Traditional light field super-resolution (LFSR) methods depend on prior knowledge of depth information. This paper presents a projection-based LFSR solution without prior information based on redefinition of the mapping function between disparity and shearing shift. Moreover, simplified variational regularization is imposed in global optimization formulation to the rendered high-resolution images. Both a synthetic dataset and a real-world dataset of light field images captured by a self-developed light field camera are used to demonstrate the state-of-the-art performance of the proposed method.

Index Terms— Light field, Super resolution, Disparity, Optimization

1. INTRODUCTION

Light field or plenoptic cameras have drawn more and more attention from industry and academia recently. These cameras are exploited on the basis of plenoptic theory which was first presented in 1992 by Adelson *et al.*[1]. Later, Levoy *et al.*[2] developed the theory and parameterized 4D light fields. In the past decade, a variety of business light field cameras have been designed such as *Lytro* and *Raytrix*. Mostly, additional optical components like lenslets are inserted between the main lens and the image sensors to capture full 4D spatio-angular information in a single photographic exposure. Although light field cameras gain huge advantages over conventional 2D cameras in many aspects such as post-capture adjustments[3][4] and one-shot depth sensing[5][6], the loss in image resolution to a small fraction of the resolution provided by the camera sensor sets bottlenecks to more practical vision tasks, such as segmentation and recognition[7][8]. To increase the spatial resolution of light field camera, the most direct way is to reduce the physical size of image sensor and enlarge the chip size. However, the approach is subject to hardware limitations and often not feasible in practice, so super resolution(SR) techniques are often focused on as Fig.1 shows.

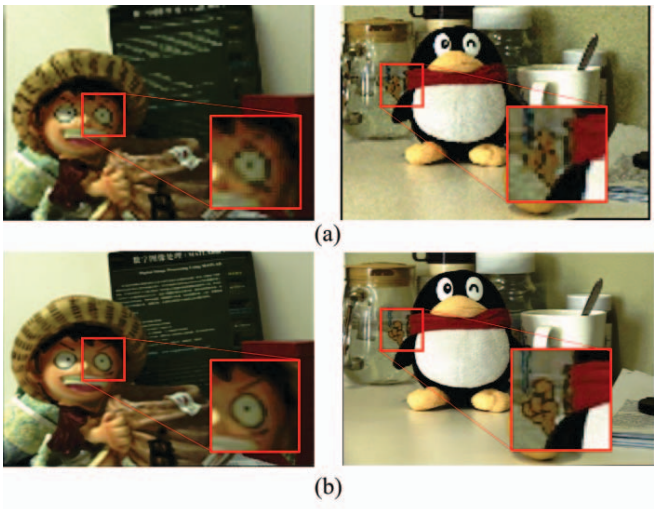

Fig. 1. Light field super resolution. (a) Low spatial-resolution images rendered from light field images. (b) All-in-focus images after super resolution based on our method.

Compared to conventional cameras, light field cameras are capable of capturing both intensity values and directions of rays from real scenes. The rays crossing the main aperture deposit on pixels according to their spatial positions and propagation directions. This property reveals essential cues for super resolution of light field cameras, *i.e.* subpixel shifted information between neighbour sub-aperture pinhole images. Lim *et al.*[9] underlined that 2D angular resolution contains spatially subpixel shifted information, which provides the redundant data used generally by SR algorithms. Georgiev *et al.*[10] also established subpixel correspondence with registration provided by the geometry of the microlens array inside the camera.

All the preceding methods for LFSR can be roughly divided into two main categories, *i.e.* variational optimization framework and projection-based algorithms. Bishop *et al.*[11] firstly designed a variational Bayesian framework to recover more information and superresolve the light field images. Similarly, after subpixel-accurate disparity estimation based on epipolar plane images(EPI), Wanner and Goldluecke[12][13] optimized a continuous variation-

al framework to generate superresolved novel views of one scene. These variational optimization frameworks depend on prior geometry information as sophisticated image priors to regularize the process, which are computational expensive. Moreover, these methods usually neglect the structure of neighboring regions. Projection-based algorithms basically rearrange light field samples with depth-variant projection. The focal stack transform introduced by Nava *et al.*[14] is an instance of projection-based methods, which could estimate the all-in-focus image of a scene at super resolution. Liang *et al.*[15] demonstrated that typical light field cameras preserved frequency components above the spatial Nyquist rate and achieved spatial resolution above lenslet resolution with projection-based algorithm. However, the method in [15] rely on internal camera parameters to obtain the shearing shift. Besides, projection-based methods don't enforce global consistency on the full resolution and are not optimized overall.

Our contributions of this paper can be summarized as follows:

(1) Redefine one mapping function between the disparity of certain pixel and its shearing shift, which relieves the dependency of internal camera parameters and depth information in the projection-based methods.

(2) Impose simplified variational regularization in global optimization formulation to the rendered high-resoltuion images, which enhances global consistency and alleviates the heavy computations for overall optimization.

2. THE PROPOSED APPROACH

The details of the proposed approach will be described in this section. The input is a raw light field image which comprises a large number of lenslet sub-images. Thus, one decoding process should be operated before analysis of super resolution. A decoding algorithm is implemented, which builds a map from the raw 2D lenslet images to the 4D light field representation *L(u,v,x,y)* under the rules of geometrical optics.

2.1. Projection-based Redefinition

Intuitively, the 4D light field *L(u,v,x,y)* can be viewed as an array of sub-aperture images that are formed by gathering the pixels of the same position in the coordinates of each microlens *(u,v)*. The sub-aperture images are equivalently captured by a pinhole camera array settled at the aperture plane. There are a variety of disparities along the sub-aperture images, which show the cues for depth of the scene. Due to the addition of depth information extracted from light field images, LFSR methods are different from those applied to conventional 2D multi-images where motion estimation and registration are usually pre-requisite. In variational framework[12][13], depth maps are required to infer geometry information and induce the warp maps between the views. As to projection-based methods in [15], when shearing

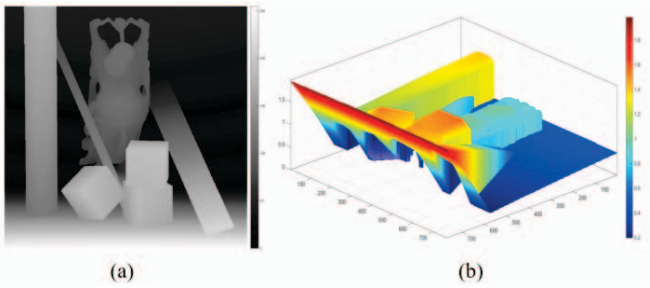

Fig. 2. An instance of sub-aperture images. (a) Extracted disparity map. (b) Consistent shearing shift.

2D light filed sample by $\boldsymbol{M_{-S}}$, its coordinate (u, x) is moved to $(u, x - Su)$. The shearing shift S is the distance between virtual plane of in-camera source light field and the lenslet array, which is relevant to camera parameters and depth information. Wanner *et al.*[16] introduced the conversion between depth and disparity as

$$d = \frac{B \times f}{\lambda} - \Delta x, \tag{1}$$

where *d* denotes disparity, B the baseline, f the focal length in pixel, λ the given depth and Δx the shift between two neighbouring images.

It is such a difficult task for depth sensing algorithms to estimate the actual depth λ without camera calibration. The internal camera parameters are also not easily known. By contrast, disparity map is naturally easier to infer from some properties of light field such as slopes of stripes on EPI.

In contrast with continuous depth, discrete disparity map is handy with facility to acquire from light fields, thus a mapping function is redefined between the disparity of certain pixel and its shifting shift as Fig.2. We follow Eq.2 from Ng *et al.*[3] to perform shearing on the full 4D data,

$$L_\alpha(u, v, x, y) = L_0(u, v, \frac{x}{\alpha} + u(1 - \frac{1}{\alpha}), \frac{y}{\alpha} + v(1 - \frac{1}{\alpha})) \tag{2}$$

when α is constant, the light fields can be refocused at certain depth, or else it is spatially variant and has linear relationship with per-sample disparity d, an all-in-focus image of the scene can be created. We focus on the latter and estimate per-sample disparity based on EPI from 4D light fields[13]. Since raw disparity map is noisy with some holes on textureless surfaces, the weighted least squared (WLS) filter is utilized to smooth it guided by aligned color image as Fig.3. Final per-sample disparity value d is ranging integer 0-255. The linear relationship between α and d is defined as

$$\alpha = M * d + N \tag{3}$$

where M and N are selected optimally through cross-validation, and consistent shearing shift S can be obtained as

$$S = 1 - \frac{1}{\alpha} \tag{4}$$

Fig. 3. Disparity refinement. (Left) Aligned color image. (Middle) EPI based Disparity Estimation. (Right) Our final disparity map.

take 2D light filed for simplicity: a) move the sample (u, x) to $(u, x - Su)$, b) project the sample to 1D spatial axis x with sheared coordinate $x - Su$ in the target image buffer. Since the buffer has higher sampling rate, or in other words subpixel accuracy, an all-in-focus image can be reconstructed with resolution higher than fixed lenslet resolution.

By analogy, spatial distribution of projected samples is similar to registration in conventional multi-frame super resolution algorithms. Next step is something like non-uniform interpolation. A pixel p in the output image O can be represented as

$$O_p = \frac{\sum_c I_c \cdot F(x_c, x_p, S_c, S_p)}{\sum_c F(x_c, x_p, S_c, S_p)} \tag{5}$$

where I denotes rearranged light field sample set, c denotes certain projected sample with spatial coordinate x_c and shearing shift S_c, x_p and S_p for the output pixel p. $F(\cdot)$ is defined as

$$F(x_c, x_p, S_c, S_p) = \frac{1 + \exp\{-|x_c - x_p|\}}{1 + \exp\{|S_c - S_p|\}} \tag{6}$$

Initial all-in-focus image of the scene has been generated with no parameters of the light field camera required during the process.

2.2. Simplified Global Optimization

The initial all-in-focus image is not optimized globally with blurred edges and low contrast. However, typical variational optimization frameworks rely on expensive deconvolution computations which are quite time-consuming. For example in Wanner's super resolution framework[13], the inclusion of warp maps and masks requires heavy calculations, which leads to the loss of efficiency.

In order to reduce the complexity of optimization, simplified variational regularization is imposed in global formulation to initial all-in-focus image Q at desired high resolution, the final optimized result R is generated by minimizing the following energy function:

$$E(R) = E_d(R) + E_g(R) \tag{7}$$

$E_d(R)$ is the data term to ensure that R does not drift too far from Q, where p is a ceratin pixel.

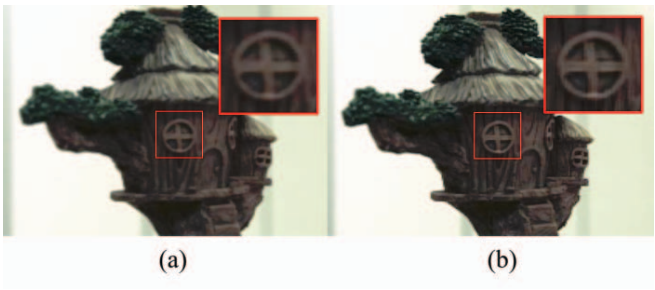

Fig. 4. Simplified Global Optimization. (a) Initial all-in-focus image by redefined projection-based method. (b) Final result after optimization.

$$E_d(R) = \sum_{p \in R} w(p)[R(p) - Q(p)]^2 \tag{8}$$

$E_g(R)$ is the variational term which acts as a regularizer mainly to enhance the contrast and global consistency,

$$E_g(R) = \sum_{p \in R} [R_x(p) - c_x Q_x(p)]^2 + [R_y(p) - c_y Q_y(p)]^2 \tag{9}$$

where R_x and R_y denote the x and y derivative of R, Q_x and Q_y of S. c_x and c_y are scalar constants that are set greater than one. In order to avoid the regions where gradient of R deviates much from that of Q, pixel-variant weighting parameter is defined as

$$w(p) = [(|R_x(p) - Q_x(p)| + 1)(|R_y(p) - Q_y(p)| + 1)]^2 \tag{10}$$

The conjugate gradient method is modified to solve the energy minimization problem in Eq.7. After optimization, the quality of all-in-focus image has been improved with more clear details as Fig.4. The whole process of optimization can be done within seconds that barely sacrifices efficiency.

3. EXPERIMENTAL RESULTS

The proposed approach can not only generate all-in-focus images at high resolution, but also super-resolved focal stack sequences refocused at different depth. Due to space limitation, we only evaluate rendered all-in-focus images at high resolution. Evaluations are performed on public or self-captured light field database with both synthetic and real-world scenes.

The HCI light field database[16] densely sampled 4D light field which contains two categories, *i.e.* synthetic scenes rendered by *Blender* and real-world objects with a *Gantry* device. Both pinhole images and ground truth depth for all views are provided at the spatial resolution of 768×768, angular resolution of 9×9. In the experiments, series of *buddha* and *mona* in *Blender* are chosen as our synthetic data. Besides, collections of light field images are captured in real-world scenes using a self-developed light field camera at the spatial resolution of 729×452, angular resolution of 9×9, the setup of experimental environment is as Fig.5.

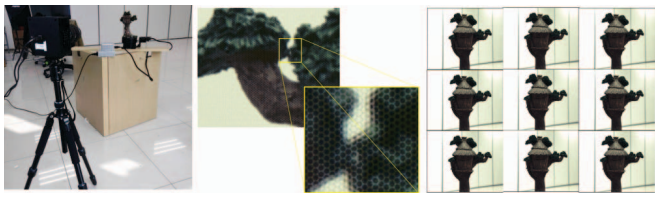

Fig. 5. Light filed images captured by a self-developed light field camera. (Left) One of our experimental setup. (Middle) Close-up of raw light field image. (Right) Decoded sup-aperture images for different views.

To validate the effectiveness of our method, synthetic light fields from HCI are preprocessed with the following steps to synthesize blurred images from original images: 1) using Gaussian filter with standard deviation 2 to smooth each sub-aperture image, 2) downsampling the sub-aperture images by a factor of 3, and 3) adding white Gaussian noise with standard deviation 0.001. Corresponding disparity maps are just downsampled accordingly. The result is compared with bicubic interpolation, simple projection-based method[15], Wanner's variational framework[13] and LFCNN[17]. Peak signal-to-noise ratio (PSNR) and the gray-scale structural similarity (SSIM)[18] are used as the performance index. The results of evaluations on the HCI dataset are shown in Table.1.

Table 1. Evaluations on the synthetic HCI dataset.

Mehtods	PSNR(dB)		SSIM	
	Buddha	Mona	Buddha	Mona
Bicubic	34.63	34.25	0.9334	0.9496
Wanner[13]	33.83	36.84	0.9107	0.9441
Projection[15]	35.38	35.43	0.9398	0.9501
Ours	**36.01**	**36.53**	**0.9451**	**0.9562**
LFCNN[17]	36.64	37.46	0.9589	0.9671

Due to degradation of sub-aperture images and corresponding disparities in preprocessing, the performance of Wanner's method drops sharply, even worse than simple bicubic interpolation. Our proposed approach can surpass the simple projection-based method by 0.63∼1.1dB for PSNR and about 0.005 for SSIM, which also achieves comparative performance with the LFCNN. Besides, the proposed method gains obvious advantages over LFCNN in the aspect that we don't need numerous light field data to train a complicated network. The algorithm is implemented in MATLAB and runs on an Intel i5 3.5GHZ and 16GB memory PC, which takes about 1∼3 minutes per light field sample, which is very competitive among these methods.

Since our light field images of real-world scenes do not have ground-truth depth or disparity, the results are just displayed for qualitative evaluation in Fig.6. Firstly, the light fields are preprocessed as the HCI synthetic data at the downsampling factor of 2, an all-in-focus image is rendered by

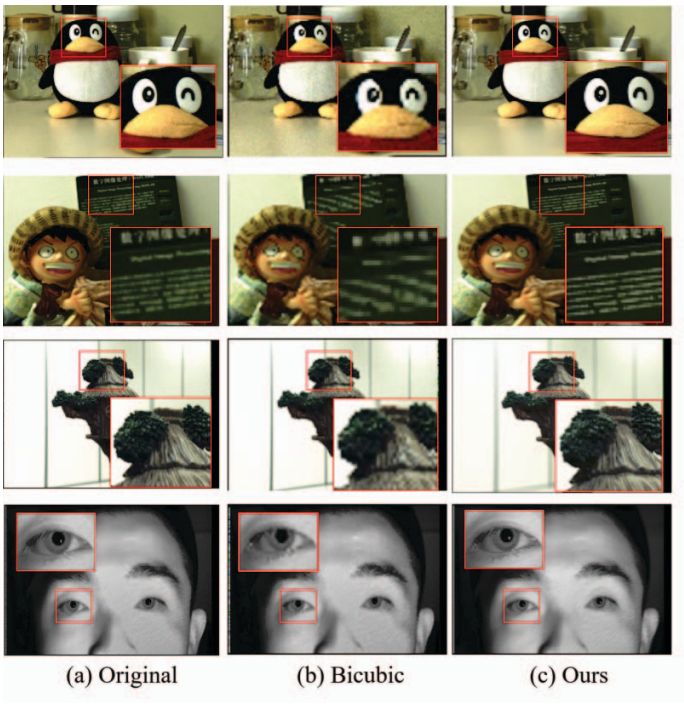

Fig. 6. Comparison of experiment results. Left column: original central viewpoints. Middle column: bicubic interpolations. Right column: our results.

refining the center-view sub-aperture image on the basis of other sub-aperture images. This all-in-focus image is upsampled again using bicubic interpolation and compared with our superresolved all-in-focus image. Our method produces better results with sharper details which are close to the original central viewpoint as Fig.6. These results demonstrate that the proposed method achieves good performance on real-world data as well.

4. CONCLUSIONS

In this paper, a simple and robust approach has been proposed to improve the spatial resolution of 4D light fields, which does not require any camera parameters or settings compared with former projection-based LFSR methods through redefinition of the mapping between disparity and shearing shift. Furthermore, simplified variational regularization has been imposed in global optimization formulation to rendered all-in-focus images at desired high resolution, which improves the visual effects with clear details. Robust super resolution results have been obtained for most scenes of light fields efficiently, especially real-world scenes captured by a self-developed light field camera.

Acknowledgement: This work is funded by National Natural Science Foundation of China Youth Fond (Grant No.61302184), National Natural Science Foundation of China Major Instrument Special Fund (Grant No. 61427811) and National Natural Science Foundation of China Fund (Grant No. 61573360).

5. REFERENCES

[1] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, no. 2, pp. 99–106, Feb. 1992.

[2] Marc Levoy and Pat Hanrahan, "Light field rendering," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 31–42.

[3] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, 2005.

[4] Liu Peng and Liu Dijun, "All-in-focus image reconstruction based on plenoptic cameras," in *Image and Graphics (ICIG), 2013 Seventh International Conference on*, July 2013, pp. 612–617.

[5] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *Computer Vision (ICCV), 2013 IEEE International Conference on*, Dec. 2013, pp. 673–680.

[6] Fei Liu, Guangqi Hou, Zhenan Sun, and Tieniu Tan, "Albedo assisted high-quality shape recovery from 4D light fields," in *Image Processing (ICIP), 2015 IEEE International Conference on*, Sept. 2015, pp. 1220–1224.

[7] Chi Zhang, Guangqi Hou, Zhenan Sun, and Tieniu Tan, "Efficient auto-refocusing of iris images for light-field cameras," in *Biometrics (IJCB), 2014 IEEE International Joint Conference on*, Sept. 2014, pp. 1–7.

[8] Shu Zhang, Guangqi Hou, and Zhenan Sun, "Eyelash removal using light field camera for iris recognition," in *Biometric Recognition*. Springer, Jan. 2014.

[9] Jae Guyn Lim, Hyun Wook Ok, Byung Kwan Park, Joo Young Kang, and SeongDeok Lee, "Improving the spatail resolution based on 4D light field data," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, Nov. 2009, pp. 1173–1176.

[10] Todor G Georgiev and Andrew Lumsdaine, "Superresolution with plenoptic 2.0 cameras," in *Signal Recovery and Synthesis*. Optical Society of America, 2009, p. S-TuA6.

[11] T. E. Bishop, S. Zanetti, and P. Favaro, "Light field superresolution," in *Computational Photography (ICCP), 2009 IEEE International Conference on*, Apr. 2009, pp. 1–9.

[12] Sven Wanner and Bastian Goldlücke, "Spatial and angular variational super-resolution of 4d light fields," in *Computer Vision–ECCV 2012*, pp. 608–621. Springer, 2012.

[13] Sven Wanner and Bastian Goldlücke, "Variational light field analysis for disparity estimation and super-resolution," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 3, pp. 606–619, Mar. 2014.

[14] F. Perez Nava and J. P. Luke, "Simultaneous estimation of super-resolved depth and all-in-focus images from a plenoptic camera," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009*, May 2009, pp. 1–4.

[15] Chia-Kai Liang and Ravi Ramamoorthi, "A light transport framework for lenslet light field cameras," *ACM Trans. Graph.*, vol. 34, no. 2, pp. 16:1–16:19, Mar. 2015.

[16] Sven Wanner, Stephan Meister, and Bastian Goldlücke, "Datasets and benchmarks for densely sampled 4d light fields," in *Annual Workshop on Vision, Modeling and Visualization: VMV*, 2013, pp. 225–226.

[17] Youngjin Yoon, Hae-Gon Jeon, Donggeun Yoo, Joon-Young Lee, and In Kweon, "Learning a deep convolutional network for light-field image super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 24–32.

[18] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.