Light Field from Micro-baseline Image Pair

Zhoutong Zhang, Yebin Liu, Qionghai Dai
Beijing Key Laboratory of Multi-dimension & Multi-scale Computational Photography (MMCP),Tsinghua University, Beijing 100084 China

Densely sampled light fields are important for many vision algorithms and applications, yet the acquisition of dense light fields is either expensive or face a trade off between spatial and angular resolution. Therefore, many methods have been devised to computationally improve the angular resolution of light fields. Those approaches can be categorized as spatial domain methods and frequency domain methods. Most spatial domain methods fall in the the category of image-based rendering (IBR), whose core task is to synthesize new views given set of images. According to [10], image-based rendering algorithms differs from their dependencies on geometry. Methods that use explicit geometry [2, 7, 9] are sensitive to inaccurate depth or disparity estimation, and high quality depth maps are hard to acquire as well. Other methods[1, 6] require little or even no geometry information, but requires a dense set of images as input. For frequency domain methods, Levin and Durand [5] leverage the dimensionality gap in the 4D frequency space of light fields to reconstruct novel views using focal stack images. Recently, Didyk et al.[3] use the phase information from a complex steerable pyramid decomposition [11] to expand view positions for 3D displays. Their method either requires multiple input images or fails to follow the basic light field structure. Noticeable ringing artifacts are easily found because of the violation against assumptions and priors.

We propose a novel phase-based light field synthesis architecture that allows high quality densely sampled light fields to be reconstructed from a micro-baseline stereo pair. The detailed work flow of our method is illustrated in Fig.1 and Fig.2. As the first step, we decompose the left view and right view using a complex steerable pyramid (Fig.1(a) and Fig.1(b)). Considering that two input images are related by a point-wise mapping(disparity), we can derive the following equation:

$$b_n^i(\mathbf{x}) = \frac{1}{2\pi}\int L(\omega)\int g_i(\mathbf{x}-\mathbf{k})e^{j\omega\cdot\mathbf{f}^{-1}(\mathbf{k})}\,d\mathbf{k}\,d\omega, \quad (1)$$

where b_n^i is the i^{th} band of the right image, $L(\omega)$denotes the Fourier Transform of the left image, f stands for the mapping function from left image to right image and f^{-1} is its inverse. The equation above can be used to synthesize novel views with sub-pixel accuracy by calculating the corresponding bands of the synthesized view with a given disparity (Fig.1(d)). This accuracy enables disparity refinement (Fig.1(f)) since the systematic error introduced in the view synthesis is relatively small compared to disparity errors. Then the phase difference between the synthesized bands and the corresponding ground truth bands is calculated (Fig.1(g)). With correct cosine fitting, the phase difference is converted to estimated disparity error and is added back to the initial estimation. Since the phase difference is noisy in nature, a filtering process [4] is required for the disparity estimation before it can be used in the new iteration. The iteration stops once the disparity improvement is lower than a threshold, and light field images are synthesized based on the optimized disparity map (Fig.1(h)).

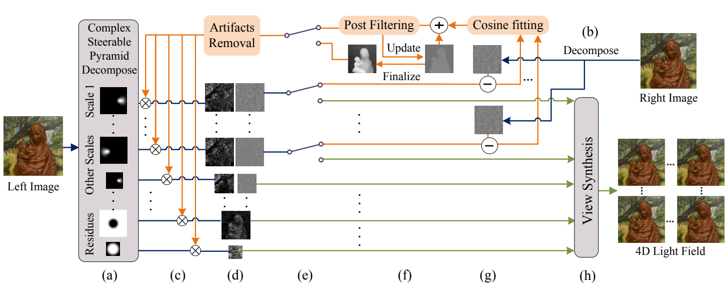

Figure 1: The pipeline of the whole proposed framework.

There are three main technical contributions of this work. First, we use a micro-baseline image pair for 4D light field reconstruction. Compared with [3], the synthesized light fields are disparity consistent. Moreover, the synthesized views are free of ringing artifacts and share the same visual quality with the input. Second, we propose a DAPS method to integrate disparity into the phase based synthesis, and achieves the best novel view synthesis quality. Our results with a simple two-image input compares well with the state of the art view synthesis method [8] using 63 views. Last, we propose an analysis by synthesis strategy to iteratively optimize the disparity based on the proposed DAPS. This solves the key problem of micro-baseline depth estimation and achieves very good results.

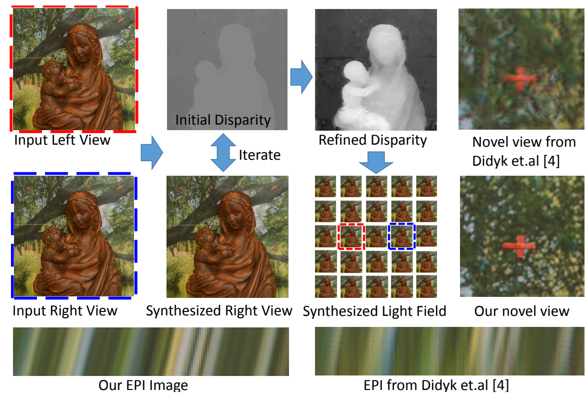

Figure 2: We reconstruct densely spaced 4D light field from a micro-baseline stereo pair. The EPI image of our reconstructed light field is free of ringing artifacts and shows more clear structures, compared with [3].

Besides the depth estimation from accidental motion shown in the paper, our work also opens the door for future analysis and processing of light fields using a phase based framework. Potential follow-up works include: a) Extending our method to more than two views, to achieve better light field depth estimation; b) The possibility of light field reconstruction from a very small number of angular views allows a more flexible trade-off between spatial and angular resolution. It thus may help improve spatial resolution, which is a key limitation of current commercial light field cameras.

[1] Chris Buehler, Michael Bosse, Leonard McMillan, Steven Gortler, and Michael Cohen. Unstructured lumigraph rendering. In *ACM SIGGRAPH*.

[2] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *SIGGRAPH*, pages 11–20, 1996.

[3] Piotr Didyk, Pitchaya Sitthi-amorn, William T. Freeman, Frédo Durand, and Wojciech Matusik. Joint view expansion and filtering for automultiscopic 3d displays. *ACM Trans. Graph.*, 32(6):221, 2013.

[4] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. In *ECCV*, 2010.

[5] Anat Levin and Fredo Durand. Linear view synthesis using a dimensionality gap light field prior. In *IEEE CVPR*, pages 1–8, 2010.

[6] Marc Levoy and Pat Hanrahan. Light field rendering. In *ACM. SIGGRAPH*, pages 31–42, 1996.

[7] William R. Mark, Leonard McMillan, and Gary Bishop. Post-rendering 3d warping. In *SI3D*, pages 7–16, 180, 1997.

[8] Sergi Pujades, Frederic Devernay, and Bastian Goldluecke. Bayesian view synthesis and image-based rendering principles. In *IEEE CVPR*, pages 3906–3913, 2014.

[9] Jonathan Shade, Steven Gortler, Li-wei He, and Richard Szeliski. Layered depth images. In *ACM SIGGRAPH*.

[10] Harry Shum and Sing Bing Kang. Review of image-based rendering techniques. In *Visual Communications and Image Processing 2000*, pages 2–13, 2000.

[11] Eero P. Simoncelli and William T. Freeman. The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *ICIP*, pages 444–447, 1995.