

FAST REALISTIC REFOCUSING FOR SPARSE LIGHT FIELDS

Chao-Tsung Huang¹, Jui Chin², Hong-Hui Chen², Yu-Wen Wang¹, Liang-Gee Chen²

¹ National Tsing Hua University, Department of Electrical Engineering

² National Taiwan University, Graduate Institute of Electronics Engineering

ABSTRACT

Digital refocusing for sparsely sampled light fields results in aliasing effect. For realistic quality, previous works performed anti-aliasing by applying time-consuming view interpolation for a heuristic number of novel views. In this paper, we study this problem by first performing a spectral analysis to give an analytical rule for the novel view number, which saves 34% of views compared to the intuitional choice. Then we propose a fast refocusing algorithm using a view interpolation method which is about 30x faster than VSRS-1D-Fast. Experimental results show the effectiveness of our approach in terms of speed and quality. For showing realistic refocusing, a light-field wafer-level-optics array is implemented, and its refocused images are compared to pictures captured by a real camera.

Index Terms— refocus, light field, view interpolation

1. INTRODUCTION

Light-field (LF) cameras can provide novel applications by manipulating the captured 4-D light fields [1]. One essential application is digital refocusing [2, 3]. For the light fields with sufficiently dense viewpoints, high-quality refocusing can be directly achieved by projection-based rendering. In contrast, for sparse light fields which can reduce storage requirement with fewer views, the direct refocusing will result in aliasing effect, and the quality is usually unacceptable as shown in Fig. 1(b). Three depth-assisted approaches can be used to reduce aliasing: depth-dependent blurring [4, 5], layered depths [6], and view interpolation [7, 8, 9]. Similar approaches can be found in computer graphics rendering [10].

The depth-dependent blurring induces discontinuous artifacts near depth boundaries as shown in Fig. 1(c) since it uses only one single view without surrounding views for occluded regions. This effect gets worse when the blur kernel (or aperture size) becomes larger. Thus this approach is only suitable for small-aperture refocusing. To overcome the boundary artifact, [6] proposed to cluster the depth map into few depth layers and apply alpha blending near depth boundaries. However, for complex scenes this method loses the details and transitions of the depth map.

Realistic refocusing quality can be achieved by interpolating novel views to compensate the insufficient view sampling as shown in Fig. 1(d). However, the computation complexity is high because lots of novel views may be needed and complex algorithms could be used, such as tri-view morphing [8], hybrid geometric-/image-based synthesis [9], and inpainting-based hole filling [7].

In this paper, we study how to reduce the computation complexity for view-interpolation-based refocusing while maintaining realistic quality. The two refocusing artifacts, aliasing effect and dis-

continuous boundary, are attacked by two primary contributions respectively: 1) A spectral analysis for digital refocusing (Section 2), which gives a rule of the minimum novel view number; 2) A fast refocusing algorithm (Section 3) which consists of 1-D line-scan view interpolation and its extension for 2-D aperture. Experimental results are shown in Section 4.

2. SPECTRAL ANALYSIS FOR VIEW-INTERPOLATION-BASED REFOCUSING

Previous works on light-field spectral analysis [11, 12] focus on how the physical world affects the captured light field. These methods rely on detailed physical information, e.g. real depth or sensor profile. Instead, our following analysis targets refocusing effect for digital images directly and thus can apply to different systems.

For non-occluded regions, view-interpolation-based refocusing is equivalent to applying a spatially-varying filter on the central view, and the ideal kernel should be a width-varying box filter. However, we will show that this is usually not the case due to fractional-pel interpolation. To meet the time-invariant assumption for Fourier analysis, we only study the localized filter response in local non-occluded regions where the disparity d_p is assumed constant.

Without loss of generality, we consider a 2-D light field with $2N + 1$ images captured at integer grid positions $v_j = -N \sim N$. The central view v_0 has the continuous signal $I_c(f)$ and discretely sampled $I_d(f) = \frac{1}{\Delta x} \sum_k I_c(f - \frac{k}{\Delta x})$ where Δx is the pixel pitch. The discrete $I_d(f)$ is used to generate the non-occluded regions of novel views at grid positions $v = -A, -A + 1/n, \dots, A$ where n is the view interpolation ratio and A is the synthetic aperture size. Note that the surrounding views are used for occluded regions but not in the following analysis. Refocusing at a focused disparity d_t is then equivalent to the following three steps:

1. **Interpolation:** For position v , derive the continuous signal $I_{c,v}(f) = I_d(f)\Delta x \cdot H(f\Delta x) \cdot e^{j2\pi v\Delta d(f\Delta x)}$ where $H(\cdot)$ is the interpolation filter and $\Delta d = d_p - d_t$.
2. **Refocusing:** Average $I_{c,v}(f)$ for the refocused $I_{c,R}(f)$. Given the Dirichlet kernel (ideal box filter) $Z_M(\nu) \triangleq \frac{\sin(\pi(2M+1)\nu)}{(2M+1)\sin(\pi\nu)}$, we have $I_{c,R}(f) = I_d(f) \cdot R_c(f\Delta x)$ where $R_c(f\Delta x) = \Delta x \cdot H(f\Delta x) \cdot Z_{nA}(\frac{\Delta d}{n}f\Delta x)$ serves as the refocusing filter in the continuous domain.
3. **Discrete Sampling:** Derive the spatially-varying digital filter $R_d(\nu) = \sum_k H(\nu - k) \cdot Z_{nN}(\frac{\Delta d}{n}(\nu - k))$ where $\nu \triangleq f\Delta x$.

The ideal $H(\cdot)$ is a rectangular function, but here we consider the practical bilinear interpolation $H(\nu) = \text{sinc}^2(x)$. As expected, $R_d(\nu)$ is an all-pass filter for $\Delta d = 0$ (infocused) and becomes a box filter for $\Delta d = n$ (exact sampling). However, for $\Delta d \neq n$, $R_d(\nu)$ will deviate from the box filter and have aliasing terms due to misalignment between $H(\nu)$ and $Z_{nN}(\frac{\Delta d}{n}\nu)$ as shown in Fig. 2.

This work was supported by the Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. MOST 103-2218-E-007-008-MY3. The authors thank Himax Technologies, Inc., for providing wafer-level optics.

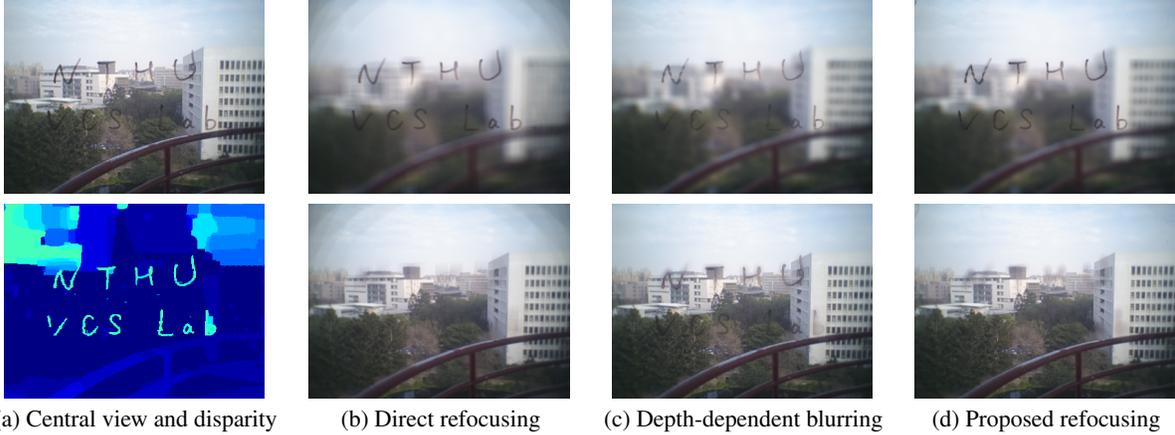


Fig. 1. Refocusing for a 5×5 sparse light field *WordsOnGlass* captured by our light-field camera. Refocused images in the top row are focused at the foreground, and those in the bottom row at the background. Direct refocusing (b) results in serious aliasing effect, and depth-dependent blurring (c) has leaking or discontinuous effect near the boundaries of words. Refocusing with view interpolation (e.g. 113 views for (d)) can solve these two artifacts, and our proposed method can maintain the realistic quality with greatly reduced computation complexity. (Readers are encouraged to zoom in the electronic paper for examining details of the refocused images.)

Interpolation Rule. For refocusing, we have to determine one single n for view interpolation. An intuitional choice is $n = \max |\Delta d|$. However, it has been shown that for somewhere $|\Delta d| < n$ we still have aliasing terms. Therefore, refocusing aliasing is inevitable for view-interpolation-based refocusing. The question becomes by how much we can accept and what criteria we should use.

In this work, we choose to use the peak aliasing magnitude for evaluation, i.e. $H(\nu = \frac{n}{|\Delta d|})$, such that the aperture size A will not affect the result. For $|\Delta d| < n$, the largest such aliasing occurs at $H(\nu = 1.4303)$, which can be considered as reasonable aliasing. Thus we suggest an interpolation rule: $H(\frac{n}{|\Delta d|}) \leq H(1.4303)$ which gives $n \geq 0.8128 \cdot |\Delta d|$ or $|\Delta d| \leq 1.2303n$.

Given a disparity map and the target d_t , we can have the suggested $n_D = \lceil 0.8128 \cdot D \rceil$ where $D = \max |\Delta d|$. For a 2-D circular aperture, our rule requires about $\pi n_D^2 A^2$ views, which saves about 34% of novel views compared to the intuitional choice $n = D$. However, hundreds of novel views are still common for large D , so we need a fast refocusing algorithm.

3. FAST REFOCUSING ALGORITHM

Our algorithm is based on two observations. First, the generated novel views will not be seen directly but summed up instead. Therefore, sophisticated view interpolation methods can be avoided to save computation. Second, interpolating views at novel 2-D positions usually involves irregular memory access and operations. Thus our algorithm is designed to have only simple 1-D line-scan operations for low complexity and regular memory access. In the following, we consider 1-D aperture first and then extend to 2-D aperture.

3.1. Line-scan view interpolation for 1-D aperture

Assume epipolar lines are parallel after rectification. Then view interpolation at horizontal grids can be performed line-by-line. Each line of right novel views $v \in (v_{j-1}, v_j]$ is interpolated from the two corresponding lines in the central view v_0 and the captured right view $v_j > 0$ as shown in Fig. 3. The case for left views $v < 0$ can be derived by symmetry.

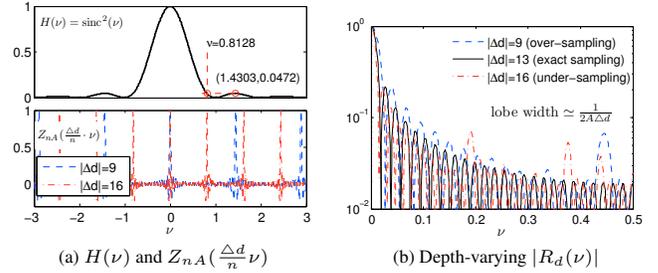


Fig. 2. Example of refocusing filter ($A = 2$ and $n = 13$). Aliasing terms appear not only for under-sampling ($|\Delta d| > n$) but also for over-sampling ($|\Delta d| < n$).

Our line scan consists of two major steps: 1) Shifting v_j for the target d_t first such that the novel views are interpolated in the positions to be refocused to avoid additional shifting and blur, which can be implicitly done by proper indexing; 2) Scanning the central pixels from right to left in one pass without Z buffer, as the i_0 to i_8 in Fig. 3, to calculate novel pixels by forward warping and hole filling.

Our second step differs from conventional view interpolation in two respects. First, the forward warping always copies pixels from the central view (red points in Fig. 3) for sharp infocused regions. Second, the hole filling simply assumes a disparity which is estimated from the closest available one on their right (green crosses in Fig. 3). This assumes the occluded region is connected to its right neighbors. When this assumption does not hold, streak artifacts could appear. However, these pixels are all in the occluded regions and are averaged for refocusing with the objects occluding them. Section 4 will show that it becomes nearly invisible after averaging.

3.2. Extension for 2-D aperture

The above line-scan method has two major benefits: simple operation and regular data access. We extend it to a right-triangular aperture while preserving these benefits. Fig. 4(a) shows the three different types of novel views. Type A is 1-D interpolated from the central

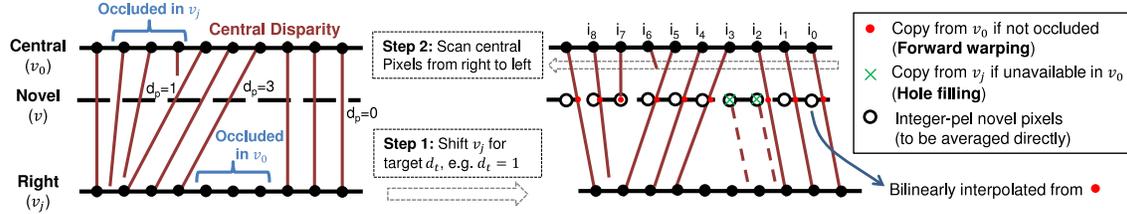


Fig. 3. 1-D line-scan view interpolation for $v \in (v_{j-1}, v_j]$ from v_0 and v_j ($v_j > 0$).

and right views. Not only pixel values but also disparities (in a similar way) are generated. Type B is 1-D interpolated from the central and top-right views, which is scanned in diagonal lines. Then Type C is generated from Types A (as the v_0) and B (as the v_j) by scanning vertical lines, which fills the triangular aperture. Finally, a full 2-D square aperture can be achieved by processing the eight symmetric triangles as shown in Fig. 4(b). Any kind of aperture shape can be formed by skipping unnecessary views.

4. EXPERIMENTAL RESULTS

Three 5×5 light fields are tested: *Tsukuba*¹ (384×288), *Truck_d2* (640×480), and *Knights_d2* (512×512). The latter two are from the Stanford light field archive² and downsampled for the test purpose with similar disparity ranges. Besides, a 5×5 wafer-level-optics (WLO) array camera is implemented for comparing the refocusing results of our algorithm to real photos. The disparity maps are derived by our multi-baseline disparity search algorithm. More refocusing results are available on our website³.

4.1. Speed and quality comparison for view interpolation

Our 1-D line-scan implementation is written in C++ and compared to the benchmark VSRS-1D-Fast implementation in 3D-HTM 10.0 [13] which is optimized for an efficient view synthesis tool. Table 1 summarizes the computation throughput of view synthesis for 1-D aperture refocusing. Since our software processes RGB format (one pixel has three samples) and VSRS-1D-Fast works on YUV 4:2:0 format, the throughput is normalized to samples/s for comparison. Though the performance varies with disparity maps, the line-scan method shows a stable speed gain from 29.9x to 37.0x.

The refocusing results are shown in Fig. 5. The line-scan method sometimes causes streak artifacts, but the refocused image shows no obvious trace due to the averaging. So, it can provide about 30x speedup with good refocusing quality compared to VSRS-1D-Fast.

4.2. Refocusing result for 2-D aperture

Fig. 6 shows the full-aperture refocusing results using the proposed algorithm. The suggested n_D reduces the refocus aliasing to the extent of no noticeable aliasing, and the proposed view interpolation delivers realistic quality. The computation times for *Tsukuba*, *Truck_d2*, and *Knights_d2* are 1.23 s (441 views), 1.54 s (113 views) and 5.85 s (797 views) respectively for a 3.4 GHz CPU. The computation mainly consists of view interpolation, image averaging, and 2-D aperture extension overhead (mostly image rotation). Especially,

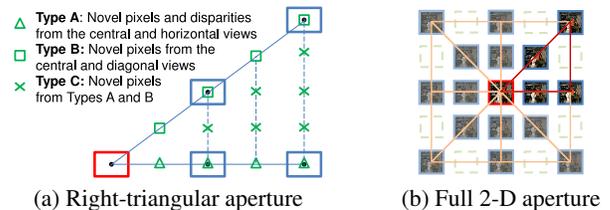


Fig. 4. Extend the 1-D line-scan method for a 2-D aperture. (a) Three different novel view types in a right-triangular aperture. (b) A full square aperture consists of eight symmetric triangles.

Table 1. Speed comparison for 1D-aperture view interpolation ($n = 8$, single thread, 3.4 GHz CPU)

| Light field (5x1) | <i>Tsukuba</i> | <i>Truck_d2</i> | <i>Knights_d2</i> |
|--------------------------------|----------------|-----------------|-------------------|
| Proposed Line-Scan (Samples/s) | 195.9M | 182.2M | 185.6M |
| VSRS-1D-Fast (Samples/s) | 5.3M | 6.1M | 5.9M |
| Throughput Ratio | 37.0x | 29.9x | 31.5x |

the line-scan view interpolation occupies 40-60% of the time after its complexity is greatly reduced by the proposed algorithm.

4.3. Experimental WLO array camera

We mount a 5×5 WLO array on a Micro Four Thirds (MFT) camera, Panasonic G5, as shown on the left of Fig. 7. The captured light fields can then be refocused after rectification and disparity estimation, and the image size is cropped to 380×284 for an equivalent field-of-view of an MFT 18-mm lens. For comparison, we also take pictures using MFT Panasonic GF2. The parameters A and d_t in our algorithm are adjusted accordingly to match the photos captured by GF2. We show one example *Car* in Fig. 7, and the refocused images show realistic and similar depth of field to the real photos. A large-aperture $f/1.88$ lens can also be simulated using $A = 2$.

5. CONCLUSION

In this paper, we show that the heavy computation problem of view-interpolation-based refocusing for sparse light fields can be alleviated. By spectral analysis, we find that the view interpolation ratio can be smaller than the conventional choice and save 34% of views. For view interpolation, we propose a fast line-scan method which can be 30x faster than VSRS-1D-Fast, and also extend it to 2-D aperture for refocusing while reserving the computation benefits. The experimental results show that our refocusing method can achieve realistic quality efficiently. Our future work will target real-time video refocusing with hardware acceleration.

¹From <http://vision.middlebury.edu/stereo/>

²<http://lightfield.stanford.edu/lfs.html>

³http://www.ee.nthu.edu.tw/chaotsung/sparse_lf_refocus



Fig. 5. 1-D aperture refocusing for the central 5×1 light field of *Tsukuba* with $D = 9.75$ and $n = 8$ (Left: refocused image; Middle: cropped image; Right: cropped novel view at $v = -11/8$). Both refocusing images look good and have no obvious artifacts. The proposed line-scan would generate novel views with streak artifacts. Some worst cases are shown here at $v = -11/8$, which get blurred in the refocused image.

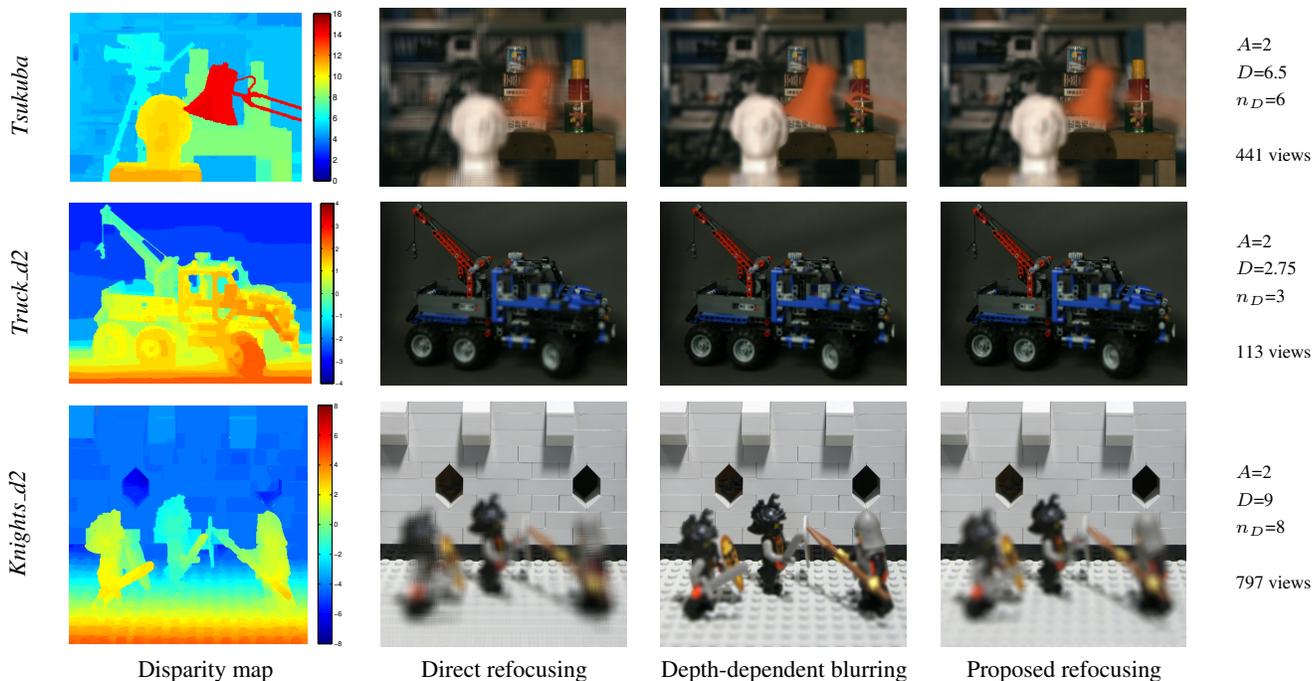


Fig. 6. 2-D refocusing results for three 5×5 light fields. *Tsukuba*: $d_p \in [4.25, 14.25]$, refocused at $d_t = 7.75$ (cans); *Truck_d2*: $d_p \in [-2.75, 2.25]$, $d_t = -0.5$ (hook); *Knights_d2*: $d_p \in [-6.5, 4.5]$, $d_t = -4.5$ (wall). The proposed refocusing uses a circular aperture. Refocus aliasing is removed with the suggested n_D , and the fast view interpolation leaves no noticeable artifacts after refocusing.

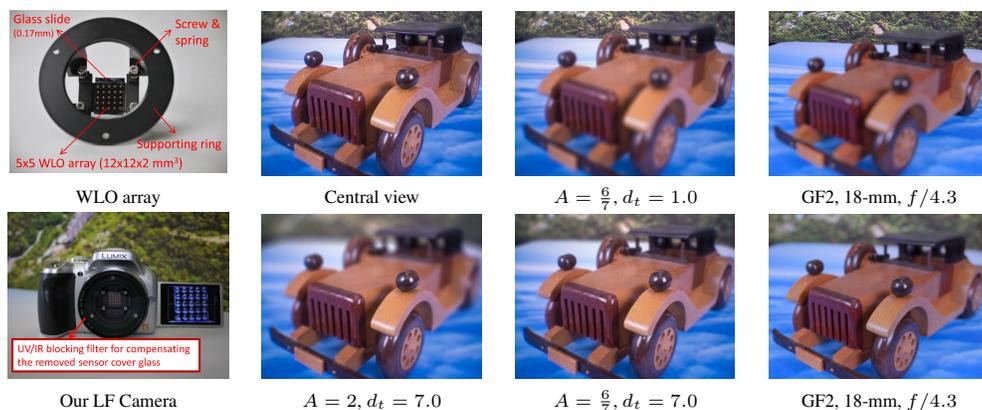


Fig. 7. Refocusing experiments on a captured light field *Car* ($d \in [1.0, 8.0]$) using WLO array camera. Two real pictures are taken by Panasonic GF2 (4th column), and we simulate two refocusing images using the corresponding aperture sizes and focal planes (3rd column) for comparison. The case for the largest synthetic aperture ($A = 2$) is also shown, which is equivalent to an MTF 18-mm $f/1.88$ lens.

6. REFERENCES

- [1] M. Harris, "Focusing on everything," *IEEE Spectrum*, vol. 49, no. 5, pp. 44–50, 2012.
- [2] Ren Ng, Marc Levoy, Mathieu Bredif, Gene Duval, Mark Horowitz, and Pat Hanrahan, "Stanford tech report ctrs 2005-02 light field photography with a hand-held plenoptic camera," 2005.
- [3] A. Isaksen, L. McMillan, and S. J. Gortler, "Dynamically reparameterized light fields," in *Proc. SIGGRAPH*, 2000, pp. 297–306.
- [4] P. Rokita, "Generating depth of-field effects in virtual reality applications," *IEEE Computer Graphics and Applications*, vol. 16, no. 2, pp. 18–21, 1996.
- [5] Yuichi Taguchi, Amit Agrawal, Ashok Veeraraghavan, Sriku-mar Ramalingam, and Ramesh Raskar, "Axial-cones: Modeling spherical catadioptric cameras for wide-angle light field rendering," *ACM Trans. Graph.*, vol. 29, no. 6, pp. 172:1–172:8, Dec. 2010.
- [6] F. Moreno-Noguer, P.N. Belhumeur, and S.K. Nayar, "Active refocusing of images and videos," *ACM Trans. Graph.*, Aug 2007.
- [7] Benjamin Huhle, Timo Schairer, Philipp Jenke, and Wolfgang Strasser, "Realistic depth blur for images with range data," in *Proceedings of the DAGM 2009 Workshop on Dynamic 3D Imaging*, 2009, pp. 84–95.
- [8] Todor Georgiev, Ke Colin Zheng, Brian Curless, David Salesin, Shree Nayar, and Chintan Intwala, "Spatio-angular resolution tradeoff in integral photography," in *Proc. Eurographics Symposium on Rendering*, 2006, pp. 263–272.
- [9] Chia-Kai Liang, Tai-Hsu Lin, Bing-Yi Wong, Chi Liu, and Homer H. Chen, "Programmable aperture photography: Multiplexed light field acquisition," in *Proc. SIGGRAPH*, 2008, pp. 55:1–55:10.
- [10] Joe Demers, "Depth of field: A survey of techniques," in *GPU Gems*, pp. 375–390. Addison-Wesley, 2004.
- [11] Jin-Xiang Chai, Xin Tong, Shing-Chow Chan, and Heung-Yeung Shum, "Plenoptic sampling," in *Proc. SIGGRAPH*, 2000, pp. 307–318.
- [12] C.-K. Liang and R. Ramamoorthi, "A light transport framework for lenslet light field cameras," *ACM Trans. Graph.*, 2015.
- [13] *Test Model 8 of 3D-HEVC and MV-HEVC*, ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 document JCT3V-H1003, Apr. 2014.