Optimized Sampling for View Interpolation in Light Fields with Overlapping Patches

D. C. Schedl and O. Bimber

Institute of Computer Graphics, Johannes Kepler University Linz
{firstname.lastname}@jku.at

Figure 1: *Comparison between two state-of-the-art view-interpolation techniques [KWR16, SBB17] with our approach. All cases upsample 64 recorded views to a total of* $15 \times 15 = 225$ *views. Applied sampling masks are shown at the top right (black: sampled, white: upsampled, red: shown view). PSNR in dB is computed for all reconstructed views with respect to the ground truth (for which all* 15×15 *views were available). Spatial and EPI close-ups are presented at the bottom, where green and red lines indicate corresponding slice positions.*

Abstract

Optimized sampling masks that reduce the complexity of camera arrays while preserving the quality of light fields captured at high directional sampling resolution are presented. We propose a new quality metric that is based on sampling-theoretic considerations, a new mask estimation approach that reduces the search space by applying regularity and symmetry constraints, and an enhanced upsampling technique using compressed sensing that supports maximal patch overlap. Our approach out-beats state-of-the-art view-interpolation techniques for light fields and does not rely on depth reconstruction.

CCS Concepts

•***Computing methodologies*** → ***Computational photography;*** *Image-based rendering;*

1. Introduction and Related Work

Sampling light fields at an adequate spatial and directional resolution is challenging. While under-sampling in the spatial domain leads to missing scene details, under-sampling results in severe bokeh artefacts in out-of-focus regions when sub-sampled directional information are combined. Camera arrays used for light-field recording suffer mainly from directional under-sampling.

Coded directional sampling and upsampling strategies for directionally sparse light-fields have shown previously to beat related view-interpolation techniques [SBB15, SBB17]. In this paper, we propose a new sampling strategy that leads to even superior results. Given an arbitrary number of samples (e.g., available cameras), we determine an optimal (by means of a proposed quality metric) configuration within a given grid of arbitrary output resolution. In contrast to previous work, our new quality metric is based on sampling-theoretical considerations. It does neither rely on learned global dictionaries or external databases for mask optimization as in [SBB17], nor on user-defined guidelines that restrict the number of mask samples as in [SBB15]. For up-sampling we apply compressed sensing that uses local dictionaries recorded with our sampling mask as in [SBB17]. Our sampling masks, however, support fully overlapping light field patches to be combined for reconstruction.

Various depth-based view interpolation techniques exist [WG14, ZLD15, PDG14, KWR16]. Depth reconstruction, however, fails for anisotropic scenes because they cannot be described sufficiently in 3D. Our approach does not require depth information. Other methods do not require depth explicitly, but still assume a Lambertian scene model [LD10, VBG17], or require a X-shaped sampling pattern [SHD*14]. We do not make any specific assumptions about the recorded scene. Recent learning-based approaches require a densely sampled input [YJY*15] or reconstruct depth [FNPS16, KWR16]. In comparison to our approach, these techniques do not find optimized sampling patterns and rely on predefined masks. Compressed sensing approaches [MWBR13, MCV14, MKU15, CC16, KHR*16] modify the optical path of classical or plenoptic cameras and use sparse bases (e.g., DCT, trained global dictionaries, or Gaussian mixture models) to reconstruct a full light field. Although we also apply compressed sensing for reconstruction, we do not rely on any precomputed bases but directly record a local dictionary with our sampling mask.

DOI: 10.2312/egs.20181034

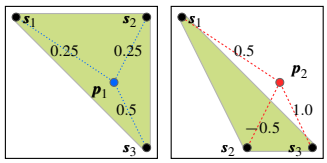

Figure 2: *Two examples of recorded samples* $\boldsymbol{S}_{1,2} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3)$ *with convex hulls (green) and position candidates* $\boldsymbol{p}_1$ *(interpolation) and* $\boldsymbol{p}_2$ *(extrapolation). Although the distances of* $\boldsymbol{p}_1$ *to* $\boldsymbol{S}_1$ *and of* $\boldsymbol{p}_2$ *to* $\boldsymbol{S}_2$ *are identical, their barycentric coordinates (overlaid numbers) vary:* $\|\boldsymbol{\lambda}\|_1$ *is* 1 *for* $\boldsymbol{p}_1$ *and* 2 *for* $\boldsymbol{p}_2$.

2. Proposed Method

We make three contributions that are presented in the following sections: First, a new quality metric that is based on sampling-theoretic considerations (Sec. 2.1). Second, a new mask estimation approach that reduces the search space by applying regularity and symmetry constrains (Sec. 2.2). Third, an enhanced upsampling technique using compressed sensing that supports maximal patch overlap (Sec. 2.3). All three contributions lead to improved upsampling results, when compared to existing techniques.

We use two-plane parametrization [LH96] and denote the angular domain by U, V and the spatial domain by S, T.

2.1. Sampling Quality Metric

A classical quality metric for sampling masks is the *maximized minimum distance (MMD)* [Kel06] which seeks for patterns that maximize the minimal Euclidean distances of each sample to its nearest neighbour. Although this ensures an even distribution of sampling positions, it does not consider the implications of each sample's contribution for interpolating or extrapolating unsampled positions.

Therefore, our new metric predicts the reconstruction quality at a position $\boldsymbol{p}$ within a given pattern of N recorded samples at positions $\boldsymbol{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N)$ in a $U \times V$ grid:

$$E_p = \sum_{i=1}^{N} d_i |\lambda_i| \,, \tag{1}$$

where d_i is the Euclidean distance between $\boldsymbol{p}$ and $\boldsymbol{s}_i$, and $\boldsymbol{\lambda}$ is the generalized barycentric coordinate of $\boldsymbol{p}$ within $\boldsymbol{S}$.

As for a MMD, a shorter d_i indicates a higher quality contribution for interpolation from close samples. However, care has to be taken for extrapolation cases. Void positions that are computed from interpolation will (for the same d_i) lead to a better reconstruction quality than positions that are computed from extrapolation.

Therefore, we additionally weight d_i by the corresponding absolute component of the generalized barycentric coordinate $|\lambda_i|$. Note, that $\|\boldsymbol{\lambda}\|_1$ is always 1 for interpolation, and greater than 1 for extrapolation (i.e., for positions $\boldsymbol{p}$ outside the convex hull of $\boldsymbol{S}$).

Thus, our metric minimizes the distances to all samples but penalize extrapolation by the absolute barycentric coordinates (cf. Fig. 2). While barycentric coordinates are uniquely defined for simple geometric cells with a small number of samples (e.g., triangles with three samples), they are equivocal for structures with an arbitrary number of samples.

In our case we solve $\boldsymbol{\lambda}$ for the *sparsest* barycentric coordinates by

$$\underset{\boldsymbol{\lambda}}{\text{minimize}} \quad \left\| \boldsymbol{p'} - \boldsymbol{S'}\boldsymbol{\lambda} \right\|_2^2, \qquad \text{subject to} \quad \|\boldsymbol{\lambda}\|_1 \leq \tau, \tag{2}$$

where $\boldsymbol{p'}$ is the extended vector $(\boldsymbol{p}, 1)^T$, $\boldsymbol{S'}$ is the extended matrix

$$\boldsymbol{S'} = \begin{pmatrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \ldots & \boldsymbol{s}_N \\ 1 & 1 & \ldots & 1 \end{pmatrix},$$

and τ is some threshold that constrains the sparsity of $\boldsymbol{\lambda}$. Equation (2) can be solved as LASSO optimization problem [Tib96].

To determine the quality of a sampling pattern that supports upsampling with full patch overlap, as explained in Sec. 2.3, we split the sampling pattern into $(U - \hat{U} - 1) \times (V - \hat{V} - 1)$ overlapping tiles of size $\hat{U} \times \hat{V}$—each (horizontally and vertically) shifted across the sampling grid at a minimal distance of 1 sample.

For each $\boldsymbol{p}$, a maximum of $\hat{U} \times \hat{V}$ quality predictions exist. We average them for tiles that support interpolation ($\|\boldsymbol{\lambda}^j\|_1 = 1$) while excluding tiles that require extrapolation ($\|\boldsymbol{\lambda}\|_1 > 1$):

$$\bar{E}_p = \langle E_p^j \rangle_j, \text{for all } j \text{ with} \|\boldsymbol{\lambda}^j\|_1 = 1. \tag{3}$$

A special case is when no interpolating tile exist for a particular $\boldsymbol{p}$. In this case, we use the minimum prediction from all available tiles.

Finally, the quality of the entire sampling mask is computed by averaging the quality metrics across all sampling positions $\boldsymbol{p}$:

$$\bar{E} = \langle \bar{E}_p \rangle_p. \tag{4}$$

2.2. Sampling Pattern Estimation

Our goal is to find the pattern of N samples within a $U \times V$ grid that minimizes $\bar{E}$. As the complexity for a brute force search is $\binom{U \times V}{N}$, the example shown in Fig. 1, with 64 samples on an 15×15 grid, leads to more than 10^{41} combinatorial possibilities.

We propose two constraints that vastly reduce the search space and therefore enable practical computation times: The sampling pattern should be *(i)* as regular and *(ii)* as symmetric as possible.

Both constraints are motivated by the idea that each sampling position in the grid is equally important. Irregularities and asymmetry, however, would lead to regions that are more densely sampled than others.

In fact, upsampling with local dictionaries, as explained in Sec. 2.3, requires a densely sampled ($\hat{U} \times \hat{V}$) region in the mask center, called *guidance area*. The guidance area serves as a basis for establishing the local dictionary used for upsampling and satisfies both our constrains. The remaining $R = N - \hat{U} \cdot \hat{V}$ samples to be distributed within the mask should do as well while minimizing $\bar{E}$.

We first compute the next highest-resolution basis grid $\tilde{U} \times \tilde{V}$ that can contain at least R samples by up-rounding:

$$\tilde{U} = \left\lceil \sqrt{\frac{U^2 R}{UV - \hat{U}\hat{V}}} \right\rceil, \qquad \tilde{V} = \left\lceil \frac{\tilde{U}V}{U} \right\rceil. \tag{5}$$

Our example in Fig. 1 ($\hat{U} = 5, \hat{V} = 5, R = 39$) results in a basis grid of 7×7.

We then scale the $\tilde{U} \times \tilde{V}$ basis grid to the $U \times V$ sampling grid. In case of a non-integer ratio of both grid resolutions, multiple permutations are possible after rounding grid positions. Resulting grid

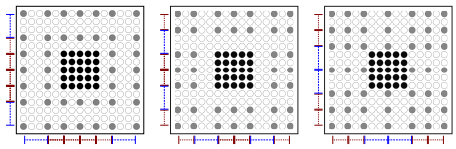

Figure 3: *The three basis grid permutations that satisfy the 4-fold rotational symmetry for our example in Fig. 1 with $N = 64$, $\tilde{U} \times \tilde{V} = 7 \times 7$ and $U \times V = 15 \times 15$ (black: guidance-area samples, gray: basis-grid samples). Varying grid spacings $\lfloor \Delta U \rfloor = \lfloor \Delta V \rfloor = 2$ and $\lceil \Delta U \rceil = \lceil \Delta V \rceil = 3$ are indicated on the left and at the bottom.*

spacings $\lfloor \Delta U \rfloor$ and $\lceil \Delta U \rceil$, where $\Delta U = \frac{U-1}{\tilde{U}-1}$, vary by a sample distance of 1 and their occurrences sum to $\tilde{U} - 1$. The spacings for the V dimension are computed analogously. To satisfy our symmetry constraint, we only pick basis grid permutations that satisfy the n-fold rotational symmetry [Wey15], where n can be 4, 2 and 1. Thus, we require a 4-fold rotational symmetry for the square configurations (as in the example shown in Fig.1), and a 2-fold rotational symmetry for non-square configurations. For our example in Fig. 1 ($\tilde{U} \times \tilde{V} = 7 \times 7$ and $U \times V = 15 \times 15$) three permutations exist (cf. Fig. 3).

Finally, we search across all basis grids for the sampling mask with a total of N samples that minimizes $\bar{E}$ by applying a stochastic search algorithm (Scatter Search [MLG06]) which removes superfluous samples with n-fold rotational symmetry outside the guidance area.

If the removal of samples breaks the n-fold rotational symmetry constraint (i.e., if R not being a multiple of n), we consider the next lower rotational symmetry case. For the example shown in Fig.1, for instance, one remaining sample position that was to be removed had neither 4-fold nor 2-fold rotational symmetric counterparts. Therefore, 1-fold rotational rotational symmetry has to be considered (i.e., no symmetry could be enforced).

2.3. Upsampling with Maximal Overlap

While the selected N samples in final mask are used for scene recording, upsampling has to be applied afterwards to determine all samples of the entire $U \times V$ grid. For upsampling we rely on a compressive sensing technique [MWBR13, SBB17] and use local dictionaries that are directly recorded by the guidance area of our sampling pattern [SBB15,SBB17]. The algorithm processes 4D light-field patches of $\hat{S} \times \hat{T} \times \hat{U} \times \hat{V}$ resolution with maximal overlap (i.e., a shift of one ray entry in both directional and spatial domains). Note, that $\hat{U} \times \hat{V}$ equals the resolution of sampling tiles described in Sec. 2.1.

Let $\boldsymbol{q'}$ be a sub-sampled light-field patch (i.e., only containing recorded ray entries). Our goal is to reconstruct an up-sampled (i.e., complete) light-field patch $\boldsymbol{q} = \boldsymbol{D\alpha}$, where $\boldsymbol{D}$ is the dictionary derived from complete patches of the guidance region (see [SBB17] for details) and $\boldsymbol{\alpha}$ the sparse coefficients found by an ADMM solver [FB15]:

$$\underset{\boldsymbol{\alpha}}{\text{minimize}} \quad \left\| \boldsymbol{q'} - \boldsymbol{\Phi D \alpha} \right\|_2^2, \qquad \text{subject to} \quad \|\boldsymbol{\alpha}\|_1 \leq \tau\,. \tag{6}$$

In Eqn. 6, $\boldsymbol{\Phi}$ is the corresponding sub-sampling matrix and τ the sparsity threshold.

After reconstructing all overlapping patches, we compute the final light field by averaging overlapping ray entries and avoiding extrapolation if possible, as explained in Sec. 2.1. Recorded ray entries remain untouched.

Scenes (N)	Marwah '13	Shi '14	Schedl '15	Kalantari '16	Schedl '17	Ours
Amethyst (64)	37.77dB	-	-	40.11dB	41.86dB	**42.08dB**
Lego (64)	28.79dB	-	-	32.87dB	35.63dB	**37.26dB**
Lego (48)	-	-	-	-	33.86dB	**35.75dB**
Cave (64)	26.51dB	-	-	30.99dB	38.57dB	**41.08dB**
Alley (64)	36.58dB	-	-	43.23dB	43.83dB	**44.35dB**
Amethyst (72)	-	36.40dB	-	-	42.18dB	**42.55dB**
Tarot (72)	-	30.19dB	-	-	37.81dB	**39.20dB**
Amethyst (69)	-	-	41.91dB	-	42.07dB	**42.43dB**
Tarot (69)	-	-	34.09dB	-	37.88dB	**39.04dB**
Tarot (48)	-	-	-	-	35.96dB	**37.54dB**
Cave (69)	-	-	29.96dB	-	39.14dB	**41.41dB**
Alley (69)	-	-	41.36dB	-	44.24dB	**45.20dB**

Table 1: *Quantitative comparison of reconstruction quality (PSNR of all reconstructed views compared to ground truth) for five scenes and five related methods [MWBR13, SHD*14, SBB15, KWR16, SBB17]. Sampling grid size was 15×15 in all cases while the number of samples (N) varied. Note, that all methods support sampling masks with arbitrary N. Cases for which the number of cameras do not support the required masks are indicated with '-'. Table 2 and Figure 4 display the applied sampling masks. Full datasets are available at dsr.files.cg.jku.at.*

3. Results and Discussion

The results presented in Tables 1,2, and in Figures 1,4 indicate that our approach out-beats state-of-the art view-interpolation techniques for light-fields and does not rely on depth reconstruction. It greatly helps to reduce the complexity of camera arrays while preserving the quality of light fields captured at high directional sampling resolution.

The key to improved upsampling results lies in the application of a guidance area to train an individual local dictionary for each recorded scene, the ability to combine fully overlapping light-field patches, and the possibility to determine optimized sampling masks with feasible computational effort. The masks are computed one-time for each camera configuration (N,U,V), and require 18 seconds to 3 minutes on an 2.7GHz Intel i5 CPU. Upsampling is significantly more time-consuming, due to the computational complexity of Eqn. 6 (which is solved on the GPU). Reconstruction time for one light field is 40 hours to 5 days on an Amazon Web Services (AWS) GPU instance (p3.2xlarge: NVIDIA Tesla V100-GPU; Intel Xeon CPU), and decreases linearly with the number of instances (e.g., down to 4-12 hours with 10 instances). However, the upsampling speed is currently the main limitation of our approach. Increasing it is our main task for future improvement.

In our experiments, we picked guidance areas that match the resolution used in state-of-the-art work [SBB17] for comparison. Finding the optimal guidance resolution can easily be achieved by repeating our approach for the small number of possible resolutions within

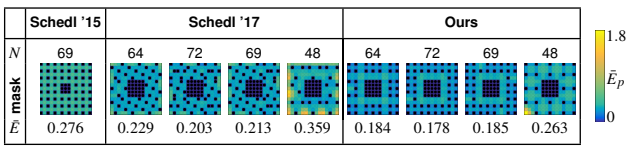

Table 2: *Comparing sampling mask qualities ($\bar{E}_p$) of approaches that apply a guidance area [SBB15, SBB17] with ours.*

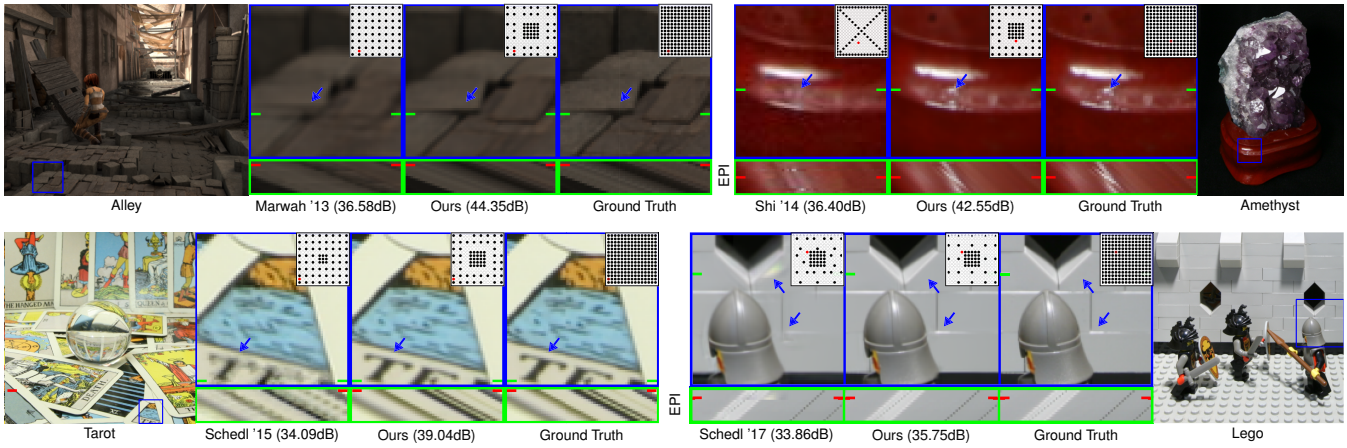

Figure 4: *Reconstruction results with [MWBR13,SHD*14,SBB15,SBB17] and with our approach, using 64, 72, 69, and 48 samples. Sampling masks are shown at the top right of spatial close-ups. EPI close-ups and corresponding slices (red and green lines) are shown at the bottom.*

the range of a given N (e.g., 3×3, 5×5, 7×7) and pick the mask with the smallest $\bar{E}$.

Interesting is to reason why (for the same N) our new sampling masks are superior to similar ones determined with the help of user-defined guidelines [SBB15] or with global dictionaries [SBB17], and what general conclusions we can make with respect to ideal patterns. With respect to Table 2, it can be seen, that all masks follow a tendency towards regularity and symmetry. In [SBB15] and in our approach this is due to applied constraints. But even for the masks learned without constrains [SBB17] this tendency can be observed (although in 45° rotated patterns). Our masks consider the influence of interpolation and extrapolation in all overlapping tiles within the mask pattern. This is the reason for gaps (depending in N extending to entire rows and columns) close to the densely sampled guidance area.

References

[CC16] CHEN J., CHAU L. P.: Light field compressed sensing over a disparity-aware dictionary. *IEEE Transactions on Circuits and Systems for Video Technology PP*, 99 (2016), 1–1. 1

[FB15] FOUGNER C., BOYD S.: Parameter selection and preconditioning for a graph form solver, 2015. 3

[FNPS16] FLYNN J., NEULANDER I., PHILBIN J., SNAVELY N.: Deepstereo: Learning to predict new views from the world's imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 5515–5524. 1

[Kel06] KELLER A.: *Myths of Computer Graphics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 217–243. 2

[KHR*16] KAMAL M. H., HESHMAT B., RASKAR R., VANDERGHEYNST P., WETZSTEIN G.: Tensor low-rank and sparse light field photography. *Computer Vision and Image Understanding 145* (2016), 172 – 181. Light Field for Computer Vision. 1

[KWR16] KALANTARI N. K., WANG T.-C., RAMAMOORTHI R.: Learning-based view synthesis for light field cameras. *ACM Trans. Graph. 35*, 6 (Nov. 2016), 193:1–193:10. 1, 3

[LD10] LEVIN A., DURAND F.: Linear view synthesis using a dimensionality gap light field prior. In *CVPR* (2010), pp. 1–8. 1

[LH96] LEVOY M., HANRAHAN P.: Light field rendering. In *SIGGRAPH* (1996), pp. 31–42. 2

[MCV14] MITRA K., COSSAIRT O., VEERARAGHAVAN A.: Can we beat hadamard multiplexing? data driven design and analysis for computational imaging systems. In *ICCP* (May 2014), pp. 1–9. 1

[MKU15] MIANDJI E., KRONANDER J., UNGER J.: Compressive image reconstruction in reduced union of subspaces. *Computer Graphics Forum 34*, 2 (2015), 33–44. 1

[MLG06] MARTÍ R., LAGUNA M., GLOVER F.: Principles of scatter search. *European Journal of Operational Research 169*, 2 (2006), 359–372. 3

[MWBR13] MARWAH K., WETZSTEIN G., BANDO Y., RASKAR R.: Compressive Light Field Photography using Overcomplete Dictionaries and Optimized Projections. *ACM Trans. Graph. 32*, 4 (2013), 1–11. 1, 3, 4

[PDG14] PUJADES S., DEVERNAY F., GOLDLUECKE B.: Bayesian view synthesis and image-based rendering principles. 1

[SBB15] SCHEDL D. C., BIRKLBAUER C., BIMBER O.: Directional super-resolution by means of coded sampling and guided upsampling. In *Computational Photography (ICCP), 2015 IEEE International Conference on* (April 2015), pp. 1–10. 1, 3, 4

[SBB17] SCHEDL D. C., BIRKLBAUER C., BIMBER O.: Optimized sampling for view interpolation in light fields using local dictionaries. *Computer Vision and Image Understanding* (2017). 1, 3, 4

[SHD*14] SHI L., HASSANIEH H., DAVIS A., KATABI D., DURAND F.: Light field reconstruction using sparsity in the continuous Fourier domain. *ACM Trans. Graph. 34*, 1 (Dec. 2014), 12:1–12:13. 1, 3, 4

[Tib96] TIBSHIRANI R.: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B 58* (1996), 267–288. 2

[VBG17] VAGHARSHAKYAN S., BREGOVIC R., GOTCHEV A.: Light field reconstruction using shearlet transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence PP*, 99 (2017), 1–1. 1

[Wey15] WEYL H.: *Symmetry*. Princeton University Press, 2015. 3

[WG14] WANNER S., GOLDLUECKE B.: Variational light field analysis for disparity estimation and super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2014). 1

[YJY*15] YOON Y., JEON H. G., YOO D., LEE J. Y., KWEON I. S.: Learning a deep convolutional network for light-field image super-resolution. In *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)* (Dec 2015), pp. 57–65. 1

[ZLD15] ZHANG Z., LIU Y., DAI Q.: Light field from micro-baseline image pair. In *CVPR* (2015), IEEE Computer Society, pp. 3800–3809. 1