# Optimized Sampling for View Interpolation in Light Fields Using Local Dictionaries

David C. Schedl, Clemens Birklbauer and Oliver Bimber

Johannes Kepler University Linz, Institute of Computer Graphics, Altenberger Strasse 69, 4040 Linz

# Abstract

We present an angular superresolution method for light fields captured with a sparse camera array. Our method uses local dictionaries extracted from a sampling mask for upsampling a sparse light field to a dense light field by applying compressed sensing reconstruction. We derive optimal sampling masks by minimizing the coherence for representative global dictionaries. The desired output perspectives and the number of available cameras can be arbitrarily specified. We show that our method yields qualitative improvements compared to previous techniques.

*Keywords:* light fields, sampling, view interpolation, superresolution, compressed sensing

# 1. Introduction and Contributions

Compared to standard digital photography, light fields offer various new options, such as refocussing, perspective changes, and 3D filtering as a postprocess. However, capturing them at an adequate resolution remains challenging. <sup>5</sup> Popular approaches typically multiplex the 4D information onto a single 2D sensor, which results either in low spatial resolution, low angular resolution, or both. Multiple sensors (e.g., a camera array) can be used to overcome the aforementioned issue. However, achieving an adequate resolution in the angular

Preprint submitted to SI: Vision, Photo, Graphics

Email address: david.schedl@jku.at (David C. Schedl)

domain requires a vast number of cameras, resulting in high construction costs 10 and complexity.

In this paper, we present an angular superresolution approach for light fields captured with sparse camera arrays. We apply compressed sensing theory for reconstruction and find optimal sampling masks for a desired number of cameras and sampling grid resolution. In contrast to related work, we avoid the need for depth reconstruction, which often fails for non-Lambertian scenes. Compressed sensing has previously been applied to light fields (e.g., in [1, 2]). One of our contributions is the use of online learned local dictionaries extracted directly from the scene sampled with an optimized mask instead of using global dictionaries that are learned offline from a set of representative pre-recorded light

- fields. Therefore, our method yields superior reconstruction results compared to related techniques. A second contribution is that, in contrast to previous work (including our own [3]), the number of samples is not constrained to the sampling pattern. Thus, our new approach allows to determine sampling masks for an arbitrary number of cameras. We compute coherence values for representative
- <sup>25</sup> global dictionaries that provide a formal basis for estimating the reconstruction quality of a given sampling pattern. We find sampling masks by minimizing the coherence. Corresponding sampling masks are optimal with respect to the representative light fields used for training the global dictionary. Our method can be applied in situations where a high angular light-field resolution is desired,
- <sup>30</sup> but camera arrays can only be constructed with a limited number of cameras (e.g., due to bandwidth limitations or high hardware costs). Our method is not suitable for light-field camera designs that do not support angular subsampling (e.g., single sensor microlens-array-based cameras).
- The remainder of the paper is organized as follows: After discussing related and previous work in Section 2, we introduce mathematical notations and revisit compressive light-field reconstruction in Section 3. Section 4 describes the proposed coherence-based quality metric, the sampling pattern optimization with offline learned global dictionaries, and the reconstruction with online learned local dictionaries. While Section 5 focusses on parameter choices and imple-

<sup>40</sup> mentation details, Section 6 is devoted to experimental results and evaluation. We conclude this article in Section 7 with a summary of limitations and future work.

# 2. Related Work

Compact light-field cameras often multiplex spatial and angular information on a single 2D sensor and thus suffer from either low spatial resolution, low angular resolution, or both. As a consequence spatial super-resolution methods for light fields have been proposed [4, 5, 6]. For camera arrays with multiple image sensors, spatial resolution is usually not an issue. However, high angular resolution requires a vast number of cameras, incurring high costs and complexity.

Angular super-resolution methods reduce the number of required cameras by reconstructing missing camera perspectives. Upsampling is applied to avoid undersampling artefacts and to enable smooth view transitions. For Lambertian scenes, depth reconstruction and subsequent view interpolation can be applied [7–8–9–10, 11–12, 13]

55 [7, 8, 9, 10, 11, 12, 13].

60

Depth reconstruction works well for adequately textured isotropic content, but can fail for more realistic scenes with non-Lambertian, anisotropic, or completely uniform objects. Non-Lambertian content cannot be described sufficiently in 3D but requires additional information, as provided in 4D light-field recordings. Thus, we compare our approach to upsampling methods that do not rely on explicit depth reconstruction.

In [14], an approach called linear view synthesis was presented that can calculate novel views from a focal stack without depth information. However, it is limited mainly to Lambertian scenes, since a focal stack covers only a

<sup>65</sup> 3D subset of a full 4D light field. The same restriction applies to the method presented in [15], where a focal stack is computed for each new perspective, and an all-in-focus image is then extracted from the focal stack.

The approach described in [16] uses a shearlet transform to reconstruct sub-

sampled epipolar-plane images of a light field, which does not require explicit

- <sup>70</sup> depth reconstruction. However, reconstruction is still based on a Lambertian scene model; the authors discussed possible extensions to non-Lambertian scenes only as part of future work. Furthermore, their sampling mask is regular, while we optimize our mask and allow arbitrary irregular patterns.
- In [17], a method specifically targeted at non-Lambertian scenes was introduced which uses sparsity in the continuous Fourier domain to reconstruct light fields from a small number of 1D viewpoint trajectories in a camera array. Although the sampling mask is sparse, the method requires very specific sampling patterns with a fixed number of cameras for capturing. In contrast, we describe how to find an optimal sampling pattern for an arbitrary number of cameras and
- <sup>80</sup> also show that we achieve higher reconstruction quality with the same number of cameras.

Recently, learning-based methods for light-field superresolution have been presented [18, 19, 13]. The approach introduced in [18], for example, trains convolutional neural networks to upsample a light field in the spatial and angular domains. However, it requires a relatively dense and regularly sampled input, while our method supports sparse and irregular samples.

85

Methods in [19, 13] use sparse input samples but rely on depth layers or depth reconstruction. In [13] two convolutional neural networks are applied one for disparity estimation, and one for view interpolation. Therefore, these methods are limited to Lambertian scenes. Furthermore, in comparison to our approach, these learning-based techniques do not optimize sampling masks, but rely on manually defined sampling patterns.

The aforementioned methods can upsample sparse light fields but require regular sampling masks. Compressed sensing approaches use irregular sampling masks to encode additional information in a low-resolution recording. The methods presented in [1, 20, 21, 22, 23, 24, 25, 26] place sampling masks in the optical path of standard cameras or compact microlens-based plenoptic cameras. Reconstructions of full light fields from the recordings are computed with sparse bases (e.g., DCT, trained global dictionaries, or Gaussian mixture models) and

- <sup>100</sup> sparsity-aware optimization methods. We also use compressed sensing theory for reconstruction, but optimize the binary angular sampling pattern of a camera array instead of using (often continuous) optical sampling masks (which affect the spatial and angular domains). Compressed sensing in the spatial domain for camera arrays was presented in [27]. Lambertian Gaussian mixture models,
- <sup>105</sup> as used in [10, 22], ignore anisotropic effects and transparencies. Corresponding methods require disparity estimations as an aditional preprocessing step. While it might be possible to reformulate the approach proposed in [22] to address the problem of choosing optimal camera sample locations, it is still limited to Lambertian scenes.
- The methods in [2] and [3] are the closest to our approach. Similarly, these techniques upsample light fields captured with a sparse camera array while avoiding depth information. Like the approach in [2], our method uses compressed sensing techniques for reconstruction. However, we extended this idea by using local dictionaries extracted from a sub-sampled light field for reconstruction. Furthermore, we present methods for computing optimal sampling
- masks for an arbitrary number of cameras and sampling grid sizes.

Our previous method [3] already presented the idea of using higher-resolution guidance areas to support up-sampling. In this article, we improved the reconstruction quality by using compressed sensing. Additionally, we present a method for computing optimal sampling configurations based on coherence val-

<sup>120</sup> method for computing optimal sampling configurations based on coherence values in a global dictionary and for an arbitrary number of cameras. In [3] we applied (empirically found) rules for estimating sampling masks that supported only specific numbers of cameras.

#### 3. Mathematical Notation and Sparse Light-Field Reconstruction

<sup>125</sup> In this section, we introduce the mathematical notations that we will use throughout this article and revisit sparse light-field reconstruction with global dictionaries (e.g., [1]).

We consider light fields captured with camera arrays and described by a regu-

lar two-plane parametrization, as discussed in [28]. Thus, rays are parametrized

- <sup>130</sup> by their intersections with two parallel planes: the camera plane UV (representing the angular domain), where the cameras are located, and the common image plane ST (representing the spatial domain), placed at a fixed distance from UV towards the objects to be captured. The indices u, v describe different camera positions on UV, and s, t address pixels in the captured perspective images  $I_{u,v}$ . We assume the light field to be regularly discretized and describe the ray intensities with the 4D matrix **L** (of size  $S \times T \times U \times V$ ) or its vectorized 1D countermant  $\mathbf{L} = uq(\mathbf{L}) = \begin{bmatrix} i & i \\ i & j \end{bmatrix}$ , which contains a sequence of
  - 1D counterpart  $\mathbf{l} = \operatorname{vec}(\mathbf{L}) = [i_{0,0}, i_{0,1}, \dots, i_{U,V}]^{\top})$ , which contains a sequence of vectorized 1D versions of the captured perspectives  $(i_{u,v} = \operatorname{vec}(I_{u,v}))$ .
- The goal of upsampling is to reconstruct a full light field **l** from its subsampled counterpart  $\mathbf{l}' = \mathbf{\Phi} \mathbf{l}$ , which only contains a subsection of all captured perspective images  $I_{u,v}$  described by the sampling matrix  $\mathbf{\Phi}$ . Since the size of  $\mathbf{l}'$ is much lower than that of **l**, this is an ill-posed, underdetermined problem. Using an approach similar to that in [1], we solve this by exploiting the compressed sensing idea. Thus, we assume that **l** is compressible and can be described as  $\mathbf{l} = \mathbf{D}\alpha$  with a sparsifying dictionary **D** and a sparse coefficient vector  $\alpha$ . The
- additional sparsity constraint for  $\alpha$  allows a robust solution to be found for  $\alpha$ , and thus also for **l**, by solving the following problem:

$$\begin{array}{ll} \underset{\alpha}{\text{minimize}} & \left\| \mathbf{l}' - \mathbf{\Phi} \mathbf{D} \alpha \right\|_{2}^{2}, \\ \text{subject to} & \left\| \alpha \right\|_{1} \leq \tau, \end{array}$$
(1)

where  $\tau$  is some threshold. This is known as the LASSO optimization problem [29]: the underdetermined system is solved by enforcing the  $\ell_1$  norm of the coefficient vector to be small, which leads to a sparse solution for  $\alpha$ . In practice, we solve the following Lagrangian formulation of the above problem using the ADMM method described in [30]:

$$\underset{\alpha}{\text{minimize}} \quad 0.5 \left\| \mathbf{l}' - \mathbf{\Phi} \mathbf{D} \alpha \right\|_{2}^{2} + \lambda \left\| \alpha \right\|_{1}.$$
(2)

Scalability for high resolutions is achieved by reconstructing the light field

patchwise. In practice, **l** contains only light-field patches of size  $(S_p \times T_p \times U_p \times V_p)$ , <sup>155</sup> which allows light fields with arbitrary resolutions to be processed.

Methods that find optimal global dictionaries  $\mathbf{D}$  for robust light-field reconstruction have been presented, for example, in [1, 2]. In these approaches, the dictionary is learned from a representative selection of light-field patches, which results in a single global dictionary that is suitable for reconstructing light fields similar to those used for learning.

160

170

In Section 4 we extend this idea by using a local dictionary extracted from the sub-sampled light field of the recorded scene. Such local dictionaries have proven to be superior to global ones, for example, in the case of image superresolution as described in [31], and also lead to better reconstruction quality in our approach.

# <sup>165</sup> 4. Global Dictionary Driven Sampling and Upsampling with Local Dictionaries

Our approach can be outlined in four consecutive steps: (i) finding optimal local sampling masks for computable camera array tile sizes based on a representative global dictionary, (ii) determining an ideal global sampling mask from the determined local tiles, (iii) recording the sparsely sampled light field with the global sampling mask, and (iv) reconstructing missing perspectives using a local dictionary being recored with the global sampling mask.

Let's assume the example for which we desire a light field with a sampling grid resolution of 15×15 perspectives being recorded with only 64 cameras (cf.

Figure 1). Finding the optimal sampling mask in a brute force search by placing 64 samples on a 15×15 grid results in  $\approx 10^{57}$  combinatoric possibilities that are infeasible to consider. Therefore, we must reduce the search space by splitting the sampling grid into Q tiles of size  $U_p \times V_p$ . In our example, we split the grid into 9 tiles of size 5×5.

Each tile can theoretically contain 1 to 25 cameras, and the total number of cameras in all tiles must match 64. Several possibilities exist to distribute cameras within a tile (e.g., 53,120 possibilities to place 5 cameras). To solve



Figure 1: Example: Finding the sampling mask for a 15x15 light field captured with only 64 cameras. We first split the target sampling grid into 9 tiles and enforce a fully sampled tile (guidance area) in the center of the mask to record the local dictionary needed for upsamplig (indicated by the blue rectangle). All remaining tiles are optimized locally such that the total number of cameras used is 64.

this problem, we search for the best pattern for each number of cameras (i.e., 1 to 25) within tile resolution using a global dictionary computed from a set of representative light fields. We explain how to achieve this in Section 4.1. Tile sampling results for common light-field patch sizes (spatial resolution x tile/directional resolution) are shown in Figure 2.

For capturing the local dictionary that is required for our upsampling, we enforce the center tile to be fully sampled and refer to this tile as *guidance area*. The remaining tiles are selected from the set of best tile patterns while optimizing a global quality metric described in more detail in Section 4.2.

A light field recorded with the resulting global sampling mask is upsampled to the full grid resolution by applying sparse reconstruction techniques with the local dictionary extracted from recording the guidance area. This is explained in more detail in Section 4.3.

# 4.1. Optimal Tile Sampling

An important factor for reconstruction quality is the incoherence of the sampling matrix  $\boldsymbol{\Phi}$  with respect to the dictionary  $\mathbf{D}$ , as stated, for example, in [1]. Random sampling matrices have proven to be a good choice in this regard; however, recent advances in the field of compressed sensing have shown that these sampling matrices can be further optimized. The restricted isometry property (RIP) is typically used to examine whether robust reconstruction of a signal from limited measurements is possible. However, as stated in [2], it does not perform well for camera arrays. Furthermore, RIP is not directly tractable. Hence, sampling matrix optimization is achieved by minimizing coherence measurements evaluated for the sub-sampled dictionary  $\mathbf{A} = \mathbf{\Phi} \mathbf{D}$ . They relate well

to the resulting reconstruction quality and are easier to compute.

Note that in our case,  $\boldsymbol{\Phi}$  represents the sampling matrix for one tile. Although we use a local dictionary for reconstruction, we apply a global dictionary, computed as in [1] and explained in Section 3, for finding optimal tile samplings. For coherence computations, the sampling matrix  $\boldsymbol{\Phi}$  and the dictionary  $\mathbf{D}$  have to be known before capturing.

Similarly to [32], we combine two different coherence formulations by a weighted sum to increase the prediction accuracy of the reconstruction quality:

$$\mu = \mu_{\text{avg}} + \beta \mu_{\text{dict}}.$$
 (3)

First, we use the average mutual coherence formulation  $(\mu_{avg})$ , as described in [33]:

$$\mu_{\text{avg}} = \frac{\left\| \tilde{\mathbf{A}}^T \tilde{\mathbf{A}} - \mathbf{I} \right\|_2^2}{K(K-1)},\tag{4}$$

215

205

where  $\hat{\mathbf{A}}$  contains normalized columns of  $\mathbf{A}$ , and K is the number of atoms (i.e., columns) in the dictionary. The average mutual coherence has been used successfully in the context of light fields (e.g., [1]). Values are small when subsampled atoms in the dictionary are different (orthogonal in the best case), which reduces ambiguities during reconstruction.

220

The second coherence value ( $\mu_{dict}$ ) in Equation (3) is the coherence difference between the original dictionary and the sub-sampled dictionary [32, 34]:

$$\mu_{\text{dict}} = \frac{\left\| \tilde{\mathbf{A}}^T \tilde{\mathbf{A}} - \tilde{\mathbf{D}}^T \tilde{\mathbf{D}} \right\|_2^2}{K(K-1)}.$$
(5)

This formulation prefers sensing matrices which result in a sub-sampled dictionary **A** with properties similar to those of **D**. Simulations in [32] showed good performance when using this coherence metric in combination with real-world images. Further details on choosing the weight  $\beta$  for Equation (3) are discussed in Section 5.

To further improve the correlation of the mean coherence value with the mean reconstruction quality, we take the number of available cameras  $N_q$  per tile into account. Thus, we extend Equation (3) as follows:

$$\tilde{\mu}_q = \mu_q \left( 1 - \frac{N_q}{U_p V_p} \right),\tag{6}$$

where  $U_p$  and  $V_p$  are the tile sizes and angular patch sizes. This improves the comparability of the coherence values for sampling matrices with different numbers of available cameras, as shown in Section 5.

230

225

With the coherence value  $\tilde{\mu}_q$ , we can now estimate the reconstruction quality for a given sampling pattern in a tile. Low coherence values indicate a high reconstruction quality.

We determine the best sampling pattern of one tile for each possible number of cameras  $j \in \{1, 2, ..., U_p V_p\}$ . Thus, we use  $\tilde{\mu}_q$  (Equation (6)) as a coherence metric and then search for the best—with respect to our representative global dictionary—sampling masks. The minimal coherence for all possible numbers of cameras (1 to  $U_p V_p$ ) is stored in a vector  $\mathbf{x} = [x_1, x_2, ..., x_{U_p V_p}]$ , where  $x_j$  is the coherence for the best tile pattern with j cameras. The best tile patterns for the tile sizes used in our experiments can be seen in Figure 2, where we computed coherence values with Equation (6) for all possible patterns and picked those with the smallest coherence.

Best tile patterns for patch sizes (i.e., spatial resolution  $\times$  tile/directional resolution) commonly used in light-field literature are shown in Figure 2.

#### 4.2. Ideal Tiling

When placing tiles in the global sampling grid, the distance to the guidance area influences the reconstruction quality. To take this distance  $d_q$  into account,

	7×7×3×3	3		<b>8×8</b>	<b>&lt;4</b> ×4					9×9×5×5	5		
000 00 1 0.371	0.213	• • • 3 0.141	0.425	2 0.225	3 0.142	4 0.112	1 0.454	2 0.210	3 0.129	4 0.099	5 0.089	6 0.083	7 0.076
0.107	● ● 5 0.083	6 0.064	5 0.104	6 0.090	7 0.081	8 0.072	8 0.072	9 0.066	10 0.062	11 0.057	12 0.053	13 0.049	14 0.045
7 0.042	8 0.021	9 0.000	9 0.062	10 0.053	11 0.044	12 0.035	15 0.041	16 0.037	17 0.033	18 0.029	19 0.024	20 0.020	21 0.016
			13 0.026	14 0.017	15 0.009	16 0.000	22 0.012	23 0.008	24 0.004	25 0.000			

Figure 2: Best tile patterns for common light-field patch sizes  $9\times9\times5\times5$ ,  $8\times8\times4\times4$ , and  $7\times7\times3\times3$ . The numbers below each pattern indicate the number of cameras  $N_q$  and the coherence  $\tilde{\mu}_q$ , respectively.

we extend Equation (6) by a weight function  $w(d_q)$ :

$$\mu'_q = \tilde{\mu}_q \, \mathbf{w}(d_q). \tag{7}$$

<sup>245</sup> We approximate the weight function with a first-degree polynomial curve as discussed in Section 5.1.

As explained earlier, we reconstruct the light field patchwise in the angular and spatial domains. Thus, we do not have a single sampling matrix  $\mathbf{\Phi}$ , but multiple ones: one  $\mathbf{\Phi}_q$  for each distinct tile of the angular sampling grid, as shown in Figure 3. To predict the reconstruction quality over the entire sampling grid, we compute a coherence value  $\mu'_q$  for each  $\mathbf{\Phi}_q$  and then average these to obtain the mean coherence value  $\bar{\mu'}$  correlating to the mean reconstruction quality. Note again that the guidance area  $(M_g)$  must be fully sampled to allow extraction of a local dictionary.

Given an angular sampling grid resolution  $U \times V$  of a light field and a number of available cameras (N), we seek to find an optimal sampling mask M such that  $\bar{\mu'}$  is minimized and thus a high reconstruction quality can be achieved. As already outlined above, the sampling mask M is split into Q tiles, where each is of size  $U_p \times V_p$ . We have already determined the best sampling patterns of one tile with corresponding coherence values for each possible number of cameras  $\mathbf{x}$  in Section 4.1. Based on  $\mathbf{x}$ , we must find the number of cameras  $N_q$  to be used for each tile (such that the sum equals N and  $\bar{\mu'}$  is minimized). We have to consider the distance weights  $w(d_q)$  of each tile in addition to the minimal coherence values stored in  $\mathbf{x}$ . Thus, we must find a solution to the following problem:

$$\begin{array}{ll} \underset{N_{1},N_{2},\ldots,N_{Q}}{\text{minimize}} & \frac{1}{Q} \sum_{q=1}^{Q} x_{N_{q}} \operatorname{w}(d_{q}), \\ \text{subject to} & \sum_{q=1}^{Q} N_{q} = N, \\ & N_{g} = U_{p} V_{p}, \\ & N_{q} \in \{1,\ldots,U_{p} V_{p}\}, \end{array}$$

$$(8)$$

where  $x_{N_q}$  is the minimal coherence  $\tilde{\mu}_q$  for the number of cameras  $N_q$ , and g is the index of the fully sampled guidance area (in the grid center).

Equation (8) can be reformulated and solved as a variant of the knapsack problem [35] by mixed-integer linear programming. The full sampling mask M is then constructed by placing the best tile sampling patterns (Secion 4.1) corresponding to the optimal number of cameras  $N_q$  determined for each tile into M.

#### 4.3. Light-Field Reconstruction with Sparsity Optimization

As stated earlier, we can apply a local dictionary for upsampling because our sampling grid contains a fully sampled tile (guidance area). The guidance area records several fully sampled light-field patches which can be used directly as atoms (i.e., columns) of our local dictionary matrix **D**. The remaining (Q - 1)tiles have different sampling matrices  $\Phi_q | q \in \{1, 2, ..., Q\}, q \notin g$ . The relation between the angular sampling mask M and one distinct sampling matrix  $\Phi_q$  is

shown in Figure 3. Since M is binary, we always sample the full spatial domain and only sub-sample the angular domain by skipping cameras in the array.



Figure 3: Relation between the angular sampling mask M (left), one tile  $M_q$  (center), and the corresponding sampling matrix  $\Phi_q$  (right). The guidance area is indicated by a blue rectangle.

Given the sampling matrices and the local dictionary, the reconstruction is performed by Equation (2). To reduce complexity and for performance reasons, tiles do not overlap in the angular domain in our case.

### <sup>285</sup> 5. Implementation and Parameterization

290

For our experiments, we used light fields from the Stanford repository [36] that are captured with a camera gantry, and synthetically rendered light fields from [37]. Eight of the Stanford light fields served as training set for the global dictionary (Figure 4), while the remaining ones were used for evaluation (Figures 7, 8, 9 and 10).

We applied patch sizes  $(S_p \times T_p \times U_p \times V_p = 9 \times 9 \times 5 \times 5, 8 \times 8 \times 4 \times 4, \text{ and } 7 \times 7 \times 3 \times 3)$ that had proved successful in previous work such as [1, 3] and permit reasonable reconstruction times.

For each patch size we learned a global dictionary by randomly selecting  $_{295}$  5,000,000 greyscale patches from the training light fields. Similarly to [1] we picked a subset of 100,000 patches with high variance from the 5,000,000 random patches. These subset patches were used to run 1,000 KSVD iterations for training a 2× overcomplete dictionary. The resulting global dictionaries were used to calculate the coherence metrics for sampling matrices.

1	ight fields	spatial resolution	disparity range
	Bracelet	320×512	-2.0 - 0.5
	Chess	400×700	-2.0 - 0.5
36]	Eucalyptus	768×640	-1.0 - 0.5
<u>1</u>	Jellybeans	256×512	-0.5 - 2.5
ainir	Bulldozer	576×768	-4.5 - 1.0
t	Truck	480×640	-1.0 - 0.5
	Bunny	512×512	-1.0 - 2.0
	Treasure	640×768	-3.0 - 1.0
36]	Amethyst	512×384	-1.0 - 1.0
test [	Lego	512×512	-2.0 - 2.0
	Tarot	512×512	-2.0 - 1.0
[37]	Cave	476×476	-3.51.0
test	Alley	400×512	-2.50.5

Table 1: Eight light fields from the Stanford repository [36] were used as training set. Three light fields from the same repository (recorded with the same camera gantry) and two additional (synthetically rendered) light fields from [37] were used for evaluation. The table shows the spatial resolution and approximate disparity ranges. The angular resolution is  $17 \times 17$ .



Figure 4: Training light fields from the Stanford repository [36]. For details on resolutions and disparities see Table 1.

#### <sup>300</sup> 5.1. Coherence Parameters and Distance Weights

We determine the optimal  $\beta$  in Equation 3 and the best w in Equation 7 experimentally. Long reconstruction times however make experimental analyses on full training light fields infeasible. Hence, we used 96 representative lightfield patches (12 per training light field) for finding optimal parameters. Repre-

- <sup>305</sup> sentative patches are patches selected for their high reconstruction complexity. To determine the reconstruction complexity of patches, we reconstructed 40,000 random patches (using randomly placed guidance areas) with different sampling masks using 3, 4 and 9 cameras, as shown in Figure 5. Note that these sampling masks contain only a single non-guidance tile. From preliminary experiments
- we know that a high number of cameras per mask and uniform camera placement improve reconstruction quality. In patches that are easy to reconstruct (e.g., uniform patches), the reconstruction quality is influenced only slightly by the sampling mask. Thus, we chose patches for which the reconstruction error (RMSE) differed most when using the 3-camera and 9-camera masks. Further-
- <sup>315</sup> more we ensured that the reconstruction error increased with decreasing number of cameras. The selected representative patches for our experiments with patch size 9×9×5×5 are shown in Figure 6 as examples. Note that for the experiments with smaller patch sizes the representative patches were cropped.



Figure 5: Different sampling masks for determining patches with high reconstruction complexity. Reconstruction quality for representative patches should decrease if the number of cameras per sampling mask is reduced.

For determining the optimal  $\beta$  in Equation 3 we sample the representative light-field patches with random tile patterns and perform reconstructions to the full patch sizes (i.e., 9×9×5×5, 8×8×4×4, and 7×7×3×3) by Equation 2. Furthermore, we compute  $\mu_{avg}$  and  $\mu_{dict}$  (Equations 4 and 5) for all random tile



Figure 6: Representative light-field patches of example patch size 9×9×5×5.

patterns. Then we search for an optimal  $\beta$  such that the reconstruction error (RMSE) and  $\tilde{\mu}_q$  of random tile patterns (Equation 6) over the exemplary patch sizes correlate the most. We use the Pearson correlation [38] as a metric for linear dependence and MATLAB's simplex search method (fminsearch) [39] to find the optimal  $\beta = 7.7$ .

We approximate the distance function (w in Equation 7) with a first-degree polynomial function. Therefore, we use the best sampling pattern of one tile <sup>330</sup> for each possible number of cameras (**x** in Section 4.1) and perform reconstructions of the representative patches. For this experiment, we are interested in the impact of the angular distance between the guidance area and the patch to reconstruct. Thus, we randomly alter the angular patch and guidance position when reconstructing representative patches with our exemplary patch

sizes. Again, we use the Pearson correlation value as a measure of correlation between RMSE and Equation 7. Using MATLAB's simplex search method, we determine  $w(d_q) = 0.0042d_q + 0.049$ . The corresponding correlations for the exemplary patch sizes  $9\times9\times5\times5$ ,  $8\times8\times4\times4$ , and  $7\times7\times3\times3$  are 0.932, 0.966 and 0.909, respectively.

340 5.2. Patch Size and Step Width

The sampling grid resolution  $(U \times V)$  is a multiple of the angular patch size in our current implementation. If multiple patch sizes are possible, we optimize  $\bar{\mu'}$ as described in Section 4.2 for each patch size and choose the sampling mask with the smaller mean coherence. The validity of this choice is shown experimentally in Table 2 for our test light fields, where we present the peak signal-to-noise ratio (PSNR) and corresponding coherence values for two different sampling grid resolutions and two possible patch sizes. These experiments indicate that larger patch sizes lead to better reconstruction results if enough cameras are available. The coherence values are in agreement with the reconstruction quality. Thus for

our results in Section 6 we used the largest applicable patch size (i.e., 9×9×5×5). For reconstruction, we allow patches to overlap in spatial domain. The overlap is defined by a spatial step width parameter that specifies the spatial distance between adjacent patches (e.g., a step width of 9 for 9×9×5×5 means no overlap) when reconstruction with Equation 2. If the step width is small more patches

<sup>355</sup> overlap, thus increasing the reconstruction quality. However, as more patches are reconstructed the computation times increases. Results in Section 6 are reconstructed with a step width of 2, which results in a good trade-off between performance and reconstruction quality.

#### 5.3. Implementation Details

We implemented our approach in Matlab and used GPU-optimizations to improve performance. In particular, we used a GPU version of the ADMM solver described in [40, 41] for reconstruction with  $\lambda = 0.01$ , relative and absolute tolerances set to 0.001 and 0.0001, and the number of iterations limited to 1000. Reconstruction time and quality depended on content, resolution and parameters. As explained in Section 4.3, we reconstruct the light field patchwise and avoid overlapping patches in the angular domain to reduce complexity;

light field size						15>	(15					
cameras $(N)$	2	7	ę.	4	ā.	6	48		4	0	3	8
patch size	9×9×5×5	7×7×3×3	9×9×5×5	7×7×3×3	9×9×5×5	7×7×3×3	9×9×5×5	7×7×3×3	9×9×5×5	7×7×3×3	9×9×5×5	7×7×3×3
Amethyst	43.42 dB	42.31 dB	42.82 dB	41.59 dB	42.39 dB	40.89 dB	41.37 dB	39.71 dB	39.30 dB	38.81 dB	36.67 dB	38.17 dB
Lego	36.74 dB	35.94 dB	36.12 dB	34.94 dB	35.33 dB	34.03 dB	33.49 dB	32.60 dB	30.80 dB	31.46 dB	26.35 dB	30.66 dB
Tarot	38.88 dB	37.17 dB	38.60 dB	36.19 dB	37.80 dB	35.19 dB	36.12 dB	33.58 dB	33.27 dB	32.37 dB	28.06 dB	31.62 dB
Cave	38.63 dB	34.22 dB	37.18 dB	32.84 dB	35.63 dB	31.54 dB	33.56 dB	29.60 dB	30.11 dB	28.29 dB	26.05 dB	27.40 dB
Alley	45.77 dB	34.77 dB	44.87 dB	32.67 dB	43.39 dB	30.92 dB	39.62 dB	28.60 dB	31.73 dB	27.26 dB	25.34 dB	26.44 dB
μ̈́	0.077	0.161	0.083	0.185	0.095	0.213	0.128	0.258	0.221	0.311	0.420	0.361

light field size						12>	12					
cameras $(N)$	9	4	20	9	4	8	4(	0	3	2	2	4
patch size	8×8×4×4	7×7×3×3	8×8×4×4	7×7×3×3	8×8×4×4	7×7×3×3	8×8×4×4	7×7×3×3	8×8×4×4	7×7×3×3	8×8×4×4	7×7×3×3
Amethyst	44.88 dB	44.26 dB	44.11 dB	43.40 dB	43.40 dB	42.23 dB	42.46 dB	41.11 dB	40.07 dB	39.37 dB	37.10 dB	38.07 dB
Lego	37.82 dB	36.76 dB	36.82 dB	35.93 dB	36.04 dB	34.92 dB	35.47 dB	33.75 dB	30.44 dB	31.80 dB	27.53 dB	30.41 dB
Tarot	40.44 dB	39.56 dB	39.53 dB	38.82 dB	38.69 dB	37.20 dB	37.80 dB	35.73 dB	34.10 dB	33.28 dB	29.18 dB	31.67 dB
Cave	42.69 dB	39.98 dB	40.75 dB	37.36 dB	38.77 dB	34.33 dB	36.49 dB	32.19 dB	30.72 dB	29.26 dB	26.19 dB	27.40 dB
Alley	47.34 dB	42.39 dB	45.64 dB	40.57 dB	44.29 dB	34.28 dB	41.97 dB	31.33 dB	30.96 dB	28.07 dB	24.98 dB	26.15 dB
μ'	0.082	0.112	0.092	0.130	0.103	0.160	0.131	0.198	0.207	0.267	0.390	0.354

Table 2: Reconstructions with varying numbers of cameras and patch sizes (multiple of light-field resolution) for our test light fields of size 15x15 and 12×12. The reconstruction quality (PSNR) correlates with the coherence  $(\mu)$ , except for minor fluctuations in the Lego scene (i.e., N = 40 and N = 32 for 15×15 and 12×12, respectively). In general, if enough cameras are available, larger patch sizes result in higher reconstruction quality. For comparability the reconstructed patches did not overlap in the spatial domain (i.e., step widths of 9, 8, and 7).

however, we allow an overlap between the patches in the spatial domain to increase reconstruction quality. The reconstructions of spatially overlapping patches are averaged in our current implementation.

370

To further increase the locality of our dictionary used for reconstruction, we computed multiple different dictionaries  $\mathbf{D}_{s,t}$  for different parts of the spatial domain. Patches were always reconstructed with the closest local dictionary. In detail, we extracted light-field patches in a sliding window fashion (spatial step width 1) from a spatial 50x50 pixel region for one dictionary, which yielded  $\sim$ 1700-2000 atoms in the dictionary (depending on the patch size). The regions 375 overlapped with a spatial step width of 25 pixels. Finally, from each atom in the dictionary its mean was subtracted, and a uniform atom was added to the dictionary for compensation. The initial solution for the sparse coefficient vector  $\alpha$  was set to reflect the mean intensity of all available rays in the sub-sampled light-field patch to speed up convergence. 380

# 6. Results

For evaluation, we used eight light fields (Figure 1) to train the global dictionary required for computing the sampling masks, and five test light fields for comparing our results with results of related approaches (see Table 1). The fully sampled light fields serve as ground truth, and we apply PSNR as a measure of 385 reconstruction quality.

We compare our approach with the three most related methods that also do not rely on depth reconstruction for upsampling: Marwah et al. 2013 [1] (we use a uniform mask for sampling and a global dictionary for upsampling), Shi et al. 2014 [17] (uses an X-shaped mask for sampling and an optimization of sparsity 390 in Fourier domain for upsampling), and our previous method [3] (uses a guidance area and uniform border as a sampling mask, and nearest-neighbor search for upsampling). The sampling grid resolution was 15×15=225 perspectives. In contrast to our current approach, the sampling masks of all related methods are constrained to specific numbers of cameras. Thus, for comparison, we chose 395

scenes	Marwah et al.	Shi et al.	Schedl et al.	Kalantari et al.	Our Method
Amethyst	39.25 dB (64)	-	-	-	43.31 dB (64)
Lego	30.25 dB (64)	-	-	-	37.08 dB (64)
Lego	30.25 dB (64)	-	-	-	34.90 dB (48)
Cave	27.96 dB (64)	-	-	-	40.02 dB (64)
Alley	38.03 dB (64)	-	-	-	45.28 dB (64)
Amethyst	-	38.09 dB (72)	-	-	43.86 dB (72)
Tarot	-	31.88 dB (72)	-	-	39.48 dB (72)
Amethyst	-	-	43.50 dB (69)	-	43.66 dB (69)
Tarot	-	-	35.68 dB (69)	-	39.47 dB (69)
Tarot	-	-	35.68 dB (69)	-	36.99 dB (48)
Cave	-	-	31.55 dB (69)	-	40.73 dB (69)
Alley	-	-	42.96 dB (69)	-	45.83 dB (69)
Lego	-	-	-	34.32 dB (64)	37.08 dB (64)
Cave	-	-	-	32.45 dB (64)	40.02 dB (64)
Alley	-	-	-	44.67 dB (64)	45.28 dB (64)

Table 3: Quantitative comparison of reconstruction quality (PSNRs) with related methods [1], [17], [3] and [13] for a sampling grid size of 15×15. Visual examples are provided for the highlighted cases in Figures 7, 8 and 9.

numbers of cameras that are applicable for the individual related methods.

Furthermore, we also compare against one exemplary state-of-the-art depthbased reconstruction method: Kalantari et al. 2016 [13] (we chose a regular sampling mask with tiles, trained the system with our eight training light fields, and used the Amethyst and Tarot scenes as test set while training).

400

405

Table 3 summarizes results and indicates that our current approach outperforms all related methods. Either, our approach delivers a better quality (higher PSNR) with the same number of cameras, or requires a lower number of cameras for the same quality (similar PSNR). Furthermore, it is much more flexible since it supports an arbitrary number of cameras for determining an optimal sampling mask.

Figures 7 and 8 present visual examples for the cases highlighted in Table 3, where the training and test light fields are similar (i.e., from the same repository [36], recorded with the same camera). Figure 9 showcases results, where the light fields drastically differ from the training set (i.e., synthetic scene and only



Figure 7: Visual comparison of reconstruction quality (PSNRs and SSIMs) with related methods [1], [17], and [3] for example cases highlighted in Table 3. Note that SSIM is given for the displayed perspective only while PSNR is computed for the entire light field. Blue rectangles mark close-ups and green rectangles present EPI images. Red dots in the sampling mask indicate the displayed view. Shown scenes are Lego, Amethyst, and Tarot.



Figure 8: Visual comparison of reconstruction quality (PSNRs and SSIMs) with a state-ofthe-art depth-based reconstruction method [13] for the Lego scene. Artefacts are visible where the depth reconstruction fails (i.e., uniform areas and occlusions). Note that SSIM is shown for the displayed perspective while PSNR is computed for the entire light field.

negative disparities).

415

Computation times for a patch size of  $9\times9\times5\times5$  and step with of 2 ranged in our implementation from 5 hours (Amethyst) to 16 hours (Tarot) on a PC with an i5-6400 CPU @ 2.70 Ghz, 24 GB RAM and a GeForce GTX 960 GPU with 4 GB RAM.

#### 7. Limitations and Future Work

We present an angular superresolution method for light fields captured with a sparse camera array. It outperforms related techniques which do not rely on depth estimation. Our main contributions are the use of a global dictionary for determining optimal sampling mask that support an arbitrary number of cameras, and the use of local dictionaries and compressed sensing theory for upsampling. Furthermore, we show that mutual coherence is a good indicator for the sampling mask quality.

Our upsampling technique can be applied to any sampling mask containing <sup>425</sup> a guidance area. However, the reconstruction quality significantly drops when non-optimized sampling masks are used (i.e., random sampling patterns, as shown in Figure 10).



Figure 9: Visual comparison of reconstruction quality (PSNRs and SSIMs) with the related methods [1] and [3]. The Cave and Alley scenes are synthetically rendered and their zerodisparity focal plane is at infinity (i.e., parallel cameras), which is quite different when compared to the light fields used for training the global dictionary. Therefore, [1] results in low quality reconstructions (e.g., blurriness), while our local dictionary generates superior results. Note that SSIM is given for the displayed perspective only while PSNR is computed for the entire light field.



Figure 10: Upsampling quality strongly degrades when non-optimized sampling patterns are used. Here we show reconstruction quality (PSNRs and SSIMs) when our upsampling technique is applied to a light field recorded with a random (i.e. non-optimized) and an optimized sampling mask (both with 49 cameras). Note that SSIM is shown for the displayed perspective while PSNR is computed for the entire light field. Reconstruction was performed with a step width of 5.

Although our experiments indicate that computed sampling masks lead to improved upsampling results even for light fields that differ much from those used in the global dictionary, a comprehensive study is required that leads to general representative global dictionaries for various applications and light field cameras. This will be part of future work.

Optimal sampling patterns for non-compressive reconstruction techniques (e.g., learning-based methods such as[13]) are on our agenda for future work.

435

430

A limiting factor is the high computational cost of finding best tile sampling patterns (as explained in Section 4.1) for tile sizes that exceed 5×5 arising from the sheer number of possible configurations. Approximate search techniques, such as scatter search and genetic algorithms, must be applied in such situations.

Furthermore, we are planning to extend our sampling mask optimization to support angular overlaps. First experiments indicate qualitative improvements while camera array sizes do not have to be a multiple of the patch size.

A main limitation is the long runtime of our implementation for upsampling. Besides general performance optimizations, integrating the disaster area detection presented in [2] is a promising option. This technique can identify regions <sup>445</sup> which benefit from a large spatial overlap during reconstruction, while other regions can be reconstructed faster without or with only minimal overlap.

While noise and vignetting is mainly an issue of single sensor light-field cameras that apply microlens arrays, it might also be a limiting factor for camera arrays with multiple sensors (e.g., in case of low-light scenes). The impact of these artefacts on our method will be investigated in future work.

#### Acknowledgements

This project was funded by the Austrian Science Fund (FWF) under contract number P 28581-N33. We thank the Stanford Computer Graphics Laboratory for providing the light fields used in this paper and the supplementary video.

#### 455 References

450

465

- K. Marwah, G. Wetzstein, Y. Bando, R. Raskar, Compressive Light Field Photography using Overcomplete Dictionaries and Optimized Projections, ACM Trans. Graph. 32 (4) (2013) 1–11.
- [2] X. Cao, Z. Geng, T. Li, Dictionary-based light field acquisition us ing sparse camera array, Opt. Express 22 (20) (2014) 24081-24095.
   doi:10.1364/OE.22.024081.
   URL http://www.opticsexpress.org/abstract.cfm?URI=

oe-22-20-24081

- [3] D. C. Schedl, C. Birklbauer, O. Bimber, Directional super-resolution by means of coded sampling and guided upsampling, in: Computational Photography (ICCP), 2015 IEEE International Conference on, 2015, pp. 1–10. doi:10.1109/ICCPH0T.2015.7168365.
  - [4] T. E. Bishop, S. Zanetti, P. Favaro, Light field superresolution, in: ICCP, 2009. doi:10.1109/ICCPH0T.2009.5559010.

- <sup>470</sup> [5] T. G. Georgiev, A. Lumsdaine, Superresolution with plenoptic 2.0 cameras, in: Signal Recovery and Synthesis, 2009. doi:10.1364/SRS.2009.STuA6. URL http://www.opticsinfobase.org/abstract.cfm?URI= SRS-2009-STuA6
- [6] V. Boominathan, K. Mitra, A. Veeraraghavan, Improving resolution and
   depth-of-field of light field cameras using a hybrid imaging system, in:
   ICCP, 2014.
  - [7] S. Wanner, B. Goldluecke, Variational light field analysis for disparity estimation and super-resolution, IEEE Transactions on Pattern Analysis and Machine Intelligence.
- 480 URL WG13\_tpami.pdf
  - [8] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, M. Gross, Scene reconstruction from high spatio-angular resolution light fields, ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH) 32 (4) (2013) 73:1– 73:12.
- [9] W. Di, L. Yebin, I. Ivo, D. Qionghai, C. Theobalt, Performance capture of high-speed motion using staggered multi-view recording, in: Pacific Graphics 2012, 2012.
  - [10] K. Mitra, A. Veeraraghavan, Light field denoising, light field superresolution and stereo camera based refocussing using a gmm light field patch
- <sup>490</sup> prior, in: CVPRW, 2012, pp. 22–28. doi:10.1109/CVPRW.2012.6239346.
  - [11] Z. Zhang, Y. Liu, Q. Dai, Light field from micro-baseline image pair., in: CVPR, IEEE Computer Society, 2015, pp. 3800-3809.
     URL http://dblp.uni-trier.de/db/conf/cvpr/cvpr2015.html# ZhangLD15
- [12] S. Pujades, F. Devernay, B. Goldluecke, Bayesian view synthesis and imagebased rendering principles, 2014.

- [13] N. K. Kalantari, T.-C. Wang, R. Ramamoorthi, Learning-based view synthesis for light field cameras, ACM Trans. Graph. 35 (6) (2016) 193:1–193:10. doi:10.1145/2980179.2980251.
- 500

URL http://doi.acm.org/10.1145/2980179.2980251

- [14] A. Levin, F. Durand, Linear view synthesis using a dimensionality gap light field prior, in: CVPR, 2010, pp. 1–8.
- [15] A. Kubota, K. Aizawa, T. Chen, Virtual view synthesis through linear processing without geometry, in: Image Processing, 2004. ICIP '04. 2004
- <sup>505</sup> International Conference on, Vol. 5, 2004, pp. 3009–3012 Vol. 5. doi: 10.1109/ICIP.2004.1421746.
  - [16] S. Vagharshakyan, R. Bregovic, A. Gotchev, Light field reconstruction using shearlet transform, IEEE Transactions on Pattern Analysis and Machine Intelligence PP (99) (2017) 1–1. doi:10.1109/TPAMI.2017.2653101.
- [17] L. Shi, H. Hassanieh, A. Davis, D. Katabi, F. Durand, Light field reconstruction using sparsity in the continuous Fourier domain, ACM Trans. Graph. 34 (1) (2014) 12:1–12:13. doi:10.1145/2682631.
   URL http://doi.acm.org/10.1145/2682631
- [18] Y. Yoon, H. G. Jeon, D. Yoo, J. Y. Lee, I. S. Kweon, Learning a deep
   <sup>515</sup> convolutional network for light-field image super-resolution, in: 2015 IEEE
   International Conference on Computer Vision Workshop (ICCVW), 2015,
   pp. 57–65. doi:10.1109/ICCVW.2015.17.
  - [19] J. Flynn, I. Neulander, J. Philbin, N. Snavely, 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.
  - [20] S. D. Babacan, R. Ansorge, M. Luessi, P. R. Mataran, R. Molina, A. K. Katsaggelos, Compressive light field sensing, IEEE Transactions on Image Processing 21 (12) (2012) 4746–4757. doi:10.1109/TIP.2012.2210237.

520

- [21] A. Ashok, M. A. Neifeld, Compressive light field imaging, Vol. 7690, 2010, pp. 76900Q-76900Q-12. doi:10.1117/12.852738.
   URL http://dx.doi.org/10.1117/12.852738
  - [22] K. Mitra, O. Cossairt, A. Veeraraghavan, Can we beat hadamard multiplexing? data driven design and analysis for computational imaging systems, in: ICCP, 2014, pp. 1–9. doi:10.1109/ICCPH0T.2014.6831800.
- 530

535

- [23] J. Chen, L. P. Chau, Light field compressed sensing over a disparity-aware
- [25] J. Chen, E. T. Chau, Egnt her compressed consing over a disparity dware dictionary, IEEE Transactions on Circuits and Systems for Video Technology PP (99) (2016) 1–1. doi:10.1109/TCSVT.2015.2513485.
- [24] E. Miandji, J. Kronander, J. Unger, Compressive image reconstruction in reduced union of subspaces, Computer Graphics Forum 34 (2) (2015) 33– 44. doi:10.1111/cgf.12539.
  URL http://dx.doi.org/10.1111/cgf.12539
- [25] M. H. Kamal, B. Heshmat, R. Raskar, P. Vandergheynst, G. Wetzstein, Tensor low-rank and sparse light field photography, Computer Vision and
- Image Understanding 145 (2016) 172 181, light Field for Computer Vision. doi:http://dx.doi.org/10.1016/j.cviu.2015.11.004. URL http://www.sciencedirect.com/science/article/pii/ S1077314215002465
  - [26] Q. YAO, K. Takahashi, T. Fujii, [paper] compressed sensing of ray space
- for free viewpoint image (fvi) generation, ITE Transactions on Media Technology and Applications 2 (1) (2014) 23-32. doi:10.3169/mta.2.23. URL http://ci.nii.ac.jp/naid/130004974701/en/
  - [27] M. H. Kamal, M. Golbabaee, P. Vandergheynst, Light field compressive sensing in camera arrays, in: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 5413–5416.
    - doi:10.1109/ICASSP.2012.6289145.

550

- [28] M. Levoy, P. Hanrahan, Light field rendering, in: SIGGRAPH, 1996, pp. 31-42. doi:http://doi.acm.org/10.1145/237170.237199.
   URL http://doi.acm.org/10.1145/237170.237199
- <sup>555</sup> [29] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society, Series B 58 (1996) 267–288.
  - [30] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (1) (2011) 1–122. doi: 10.1561/2200000016.
    - URL http://dx.doi.org/10.1561/2200000016

560

- [31] G. Freedman, R. Fattal, Image and video upscaling from local selfexamples, ACM Trans. Graph. 28 (3) (2010) 1–10. doi:http://doi.acm. org/10.1145/1531326.1531328.
- <sup>565</sup> [32] H. Bai, G. Li, S. Li, Q. Li, Q. Jiang, L. Chang, Alternating optimization of sensing matrix and sparsifying dictionary for compressed sensing, IEEE Transactions on Signal Processing 63 (6) (2015) 1581–1594. doi:10.1109/ TSP.2015.2399864.
- [33] J. M. Duarte-Carvajalino, G. Sapiro, Learning to sense sparse signals: Si multaneous sensing matrix and sparsifying dictionary optimization, IEEE
   Transactions on Image Processing 18 (7) (2009) 1395–1408. doi:10.1109/
   TIP.2009.2022459.
  - [34] N. Cleju, Optimized projections for compressed sensing via rankconstrained nearest correlation matrix, CoRR abs/1309.3676.
- 575 URL http://arxiv.org/abs/1309.3676
  - [35] G. B. Dantzig, Discrete-variable extremum problems, Operations research 5 (2) (1957) 266–288.
  - [36] V. Vaish, A. Adams, The (new) stanford light field archive, http:/lightfield. stanford. edu.

- 580 [37] T. Roosendaal, Sintel. blender foundation, durian open movie project (2011). URL https://durian.blender.org/
  - [38] R. A. Fisher, Statistical methods for research workers, Genesis Publishing Pvt Ltd, 1925.
- <sup>585</sup> [39] J. C. Lagarias, J. A. Reeds, M. H. Wright, P. E. Wright, Convergence properties of the nelder-mead simplex method in low dimensions, SIAM Journal on optimization 9 (1) (1998) 112–147.
  - [40] N. Parikh, S. Boyd, Block splitting for distributed optimization, Mathematical Programming Computation 6 (1) (2014) 77–102. doi:10.1007/s12532-013-0061-8

590

s12532-013-0061-8. URL http://dx.doi.org/10.1007/s12532-013-0061-8

[41] C. Fougner, S. Boyd, Parameter selection and pre-conditioning for a graph form solver (2015). arXiv:arXiv:1503.08366.