

# Automatika

Journal for Control, Measurement, Electronics, Computing and Communications



ISSN: 0005-1144 (Print) 1848-3380 (Online) Journal homepage: <https://www.tandfonline.com/loi/taut20>

## Overcoming spatio-angular trade-off in light field acquisition using compressive sensing

Nicol Dlab, Filip Cimermančić, Ivan Ralašić & Damir Seršić

To cite this article: Nicol Dlab, Filip Cimermančić, Ivan Ralašić & Damir Seršić (2020) Overcoming spatio-angular trade-off in light field acquisition using compressive sensing, *Automatika*, 61:2, 250-259, DOI: [10.1080/00051144.2020.1715582](https://doi.org/10.1080/00051144.2020.1715582)

To link to this article: <https://doi.org/10.1080/00051144.2020.1715582>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 21 Jan 2020.



Submit your article to this journal [↗](#)



Article views: 242



View related articles [↗](#)



View Crossmark data [↗](#)



# Overcoming spatio-angular trade-off in light field acquisition using compressive sensing

Nicol Dlab, Filip Cimermančić, Ivan Ralašić and Damir Seršić

Department of Electronic Systems and Information Processing, University of Zagreb Faculty of Electrical Engineering and Computing, Zagreb, Croatia

## ABSTRACT

In contrast to conventional cameras which capture a 2D projection of a 3D scene by integrating the angular domain, light field cameras preserve the angular information of individual light rays by capturing a 4D light field of a scene. On the one hand, light field photography enables powerful post-capture capabilities such as refocusing, virtual aperture, depth sensing and perspective shift. On the other hand, it has several drawbacks, namely, high-dimensionality of the captured light fields and a fundamental trade-off between spatial and angular resolution in the camera design. In this paper, we propose a compressive sensing approach to light field acquisition from a sub-Nyquist number of samples. Using an off-the-shelf measurement setup consisting of a digital projector and a Lytro Illum light field camera, we demonstrate the efficiency of the compressive sensing approach by improving the spatial resolution of the acquired light field. This paper presents a proof of concept with a simplified 3D scene as the scene of interest. Results obtained by the proposed method show significant improvement in the spatial resolution of the light field as well as preserved post-capture capabilities.

## ARTICLE HISTORY

Received 26 July 2019  
Accepted 3 January 2020

## KEYWORDS

Compressive sensing (CS);  
light field; camera; projector;  
measurement setup; sparse  
optimization

## 1. Introduction

Since the invention of the digital camera in mid-1970s, lots of effort has been given into improving the digital photography. Most of the research in this field focused on increasing the spatial resolution and reducing the physical size of the imaging devices. Conventional cameras capture two-dimensional projections of the three-dimensional world with high spatial resolution. Even though the spatial information is well preserved, the information contained in other dimensions of the light field is lost in the acquisition process.

Plenoptic function  $L(x, y, z, \theta, \phi, \lambda, t)$  is a 7D function which has been introduced in [1]. It represents the radiance of the light rays emitted by a scene and received by an observer at every point  $(x, y, z)$  in the 3D space, along any direction  $(\theta, \phi)$ , at any time instance  $(t)$  and for all wave lengths  $(\lambda)$ . Plenoptic function contains detailed information about the light field within the observed scene.

The idea of light field photography was introduced in the pioneer work of Lippman [2]. The main goal of light field photography is to record not only the scalar intensity of the observed scene, but also the angular information of the light arriving at each pixel of the imaging sensor. Angular information enables novel post-capture capabilities such as refocusing, depth sensing, perspective shifts and synthetic apertures. Unfortunately, that

comes with an inherent trade-off in the spatio-angular resolution of the light field cameras, resulting in the fact that the spatial resolution of the light field images does not meet the standards of the modern digital photography.

In recent years, light field photography became a popular research area and multiple methods for light field acquisition and processing have been proposed. Some of the methods focus exclusively on the light field acquisition, while the others work on improving different aspects of the captured light fields using advanced post-processing. Light field acquisition using temporal multiplexing is a simple idea where a static scene is observed sequentially from multiple viewpoints using a conventional camera. In [3], the authors used a camera gantry in order to acquire multiple two-dimensional images. A 4D light field is then reconstructed from multiple 2D projections of the observed scene. In [4], a prototype of a programmable aperture light field camera was proposed. Programmable aperture enables multiplexed light field acquisition using a conventional camera. Compressive light field camera based on a programmable aperture was proposed in [5], where the authors introduced novel optical designs along with improved reconstruction algorithms in order to obtain light field from a single camera exposure. Spatial multiplexing using camera arrays is another way to acquire

light fields. In [6], the authors built a camera array which consists of  $10 \times 10$  cameras. This setup enabled light field capture with high spatial resolution, demonstrating the ability to see through partial occlusions. However, this setup needs careful calibration and is too bulky for practical use. Ng [7,8] prototyped a portable light field camera by inserting a micro-lens array (MLA) between the sensor and the main lens. Each micro-lens enables capturing information on both the intensity and the direction of a light ray. Later, this prototype resulted in the first consumer grade light field camera produced by Lytro, followed by an improved camera model named Lytro Illum. A major drawback of the MLA-based light field acquisition is the low spatial resolution of the captured images, which is due to the fact that a single imaging sensor is shared to capture both the spatial and the angular information.

There are two main approaches for improving the spatial and angular resolution of the light field. Hybrid light field imaging systems like the ones proposed in [9,10] consist of a conventional high-resolution camera and a light field camera. In such imaging systems, resolution of the individual light field sub-aperture is enhanced using the information from the conventional camera image. Recently, deep learning methods for light field super-resolution were proposed [11–13]. These methods learn underlying statistical distribution of a training dataset consisting of different light field examples. After the training, such methods are able to ‘hallucinate’ higher resolution details in the angular and spatial domain of the light fields. Contrary to the previously mentioned hybrid methods, deep learning methods for light field super-resolution are not based on physical measurements.

In this paper, we present a compressive sensing (CS) approach to the light field acquisition. We design and implement an off-the-shelf setup consisting of a digital projector and a Lytro Illum light field camera. The proposed setup enables improvement of the spatial resolution of the captured light field and overcomes the aforementioned spatio-angular trade-off. In contrast with previously mentioned hybrid imaging systems, our measurement system acquires high-resolution images of the observed scene from a sub-Nyquist number of samples. The results presented in this paper show significant improvement in the spatial resolution of the sampled light field as well as preserved post-capture capabilities.

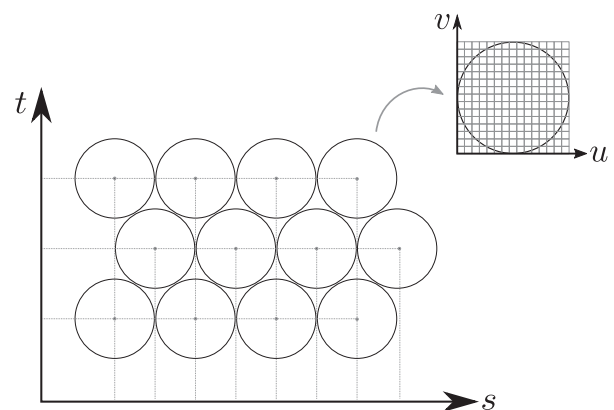
The remainder of this paper is organized as follows: Section 2 describes the light field parameterization and acquisition, Section 3 offers a short introduction into compressive sensing framework, while Section 4 describes our measurement setup along with the acquisition and reconstruction process. Experimental results are presented in Section 5, while conclusions are drawn in Section 6.

## 2. Light field parameterization and acquisition

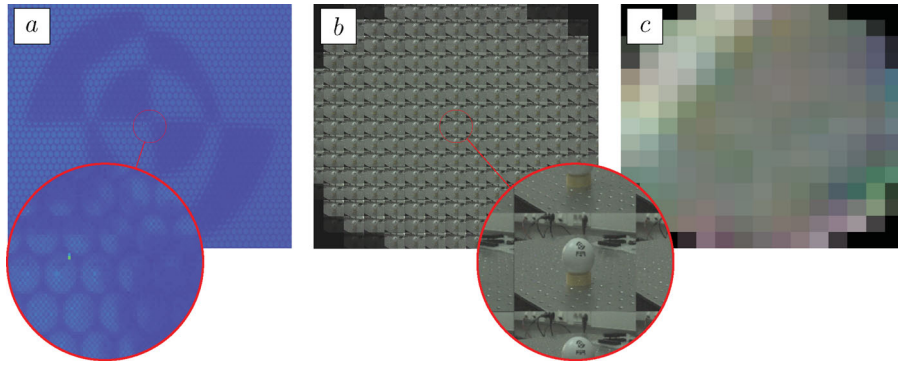
Plenoptic 7D function can be simplified to a 5D function by observing the light field at a single time instance (i.e. omitting  $t$ ) and by adding a colour filter array on the photo-sensor (i.e. omitting  $\lambda$ ). Moreover, by assuming propagation of light ray in a space without occlusions, the plenoptic function can be further simplified to a 4D light field representation  $LF(u, v, s, t)$  that was introduced in [3].  $(s, t)$  denote the spatial, while  $(u, v)$  are the angular dimensions of the light field. In this paper, we use the two-plane parameterization of the light field.

Visualization of the two-plane light field parameterization is shown in Figure 1. Lytro Illum camera contains a hexagonal micro-lens array that is placed in front of the imaging sensor. Each micro-lens is positioned at its corresponding coordinate in the  $(s, t)$  plane. Number of lenslets defines the spatial resolution of the captured light field. Imaging sensor is placed at the focal distance of the MLA. Pixels of the imaging sensor lie in the  $(u, v)$  plane. Number of pixels behind each of the micro-lens defines the angular resolution of the light field. In Figure 2(a), an inset from the raw Lytro Illum camera image is shown, and the hexagonal arrangement of micro-lenses is visible in the magnified detail.

Figure 2(b) shows the tiled light field format. Each of the sub-images corresponds to a single sub-aperture of the Lytro camera. Sub-aperture images are constructed by taking a single pixel with specific  $(u, v)$  under each micro-lens. Essentially, such an extraction corresponds to holding  $(u, v)$  fixed and considering all  $(s, t)$  coordinates in the 4D light field representation.



**Figure 1.** Visualization of the two-plane light field parameterization.  $(s, t)$  coordinate system corresponds to the spatial information of the light field. Each circle corresponds to a single micro-lens and the number of micro-lenses defines the spatial resolution of the light field.  $(u, v)$  coordinate system corresponds to the angular information of the light field. A single lenslet is shown with its corresponding pixels of the imaging sensor. Number of sensor pixels behind the micro-lens defines the angular resolution of the light field.



**Figure 2.** (a) Raw image from the Lytro Illum camera sensor with hexagonally arranged micro-lens array on top of an imaging sensor with Bayer colour filter array, (b) tiled view of the captured light field. Each sub-image corresponds to a different viewpoint. (c)  $15 \times 15$  pixel image under a single micro-lens. Number of pixels defines the angular resolution of the light field.

Figure 2(c) shows a single micro-lens image consisting of  $15 \times 15$  pixels, which corresponds to the angular resolution of the Lytro Illum camera. Notice the significant vignetting of the boundary pixels of the micro-lens.

Light field imaging enables a wide variety of post-capture capabilities such as digital refocusing, viewpoint change, virtual aperture and depth sensing [7,8,14]. In order to easily model the post-capture manipulation of the light field, synthetic light field  $L'(u', v', s', t')$  was introduced in [7] inspired by the synthetic sensor plane equation  $E(s', t')$  given by:

$$E(s', t') = \frac{1}{D^2} \iint L'(u', v', s', t') A(u', v') \cos^4 \theta \, du \, dv, \quad (1)$$

where  $D$  is separation between the sensor and aperture,  $A$  is the aperture function, and  $\theta$  is the angle of incidence that ray  $(u', v', s', t')$  makes with the film plane [7]. All the post-capture manipulations of the light field are essentially methods that numerically approximate this integral. For example, refocusing is conceptually just a summation of shifted versions of the images that form through pinholes over the entire  $(u, v)$  aperture.

The main drawback of most light field imaging devices is the aforementioned spatio-angular trade-off. Specifically, Lytro Illum camera contains a 40MP imaging sensor. Spatial resolution of the light field captured using the Lytro Illum camera is limited by the number of the micro-lenses placed above the imaging sensor and equals around 0.3MP after decoding it into the 4D light field format. Notice that this resolution is rather small compared to the spatial resolution of the modern imaging sensors.

In this paper, our goal is to improve the spatial resolution of the light fields while retaining the original angular resolution by using a simple off-the-shelf CS measurement setup and by applying compressive sensing principles. In the following section, we provide a short overview of the CS framework.

### 3. Compressive sensing

Compressive sensing is a signal processing framework which consists of a linear measurement and a nonlinear reconstruction process which is based on sparse optimization [15–17]. Research on various applications of the CS framework [18–22] has been very active in the recent years in different scientific areas.

The CS measurement process can be modelled using inner products between the original signal  $\mathbf{x}$  and a collection of measurement vectors  $\phi_i$  as in  $y_i = \langle \mathbf{x}, \phi_i \rangle$ , where  $i = 1 \dots M$ . Original signal  $\mathbf{x}$  is  $N$ -dimensional, while the CS measurements are  $M$ -dimensional. In the CS measurement process, only a subset of measurements defined by the Shannon–Nyquist theorem is acquired. Since  $M \ll N$ , the quotient  $r = M/N$  is called the subsampling ratio.

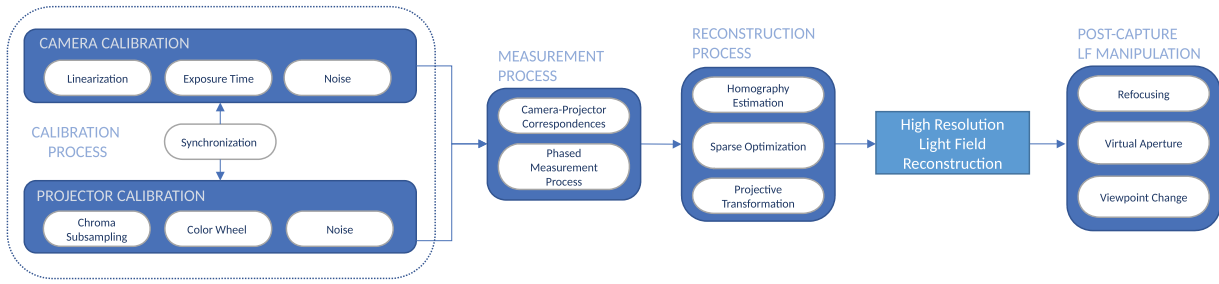
Measurement results  $y_i$  can be arranged into an  $M \times 1$  vector  $\mathbf{y}$  and the measurement vectors  $\phi_i^T$  can be written as rows of an  $M \times N$  matrix  $\Phi$ . CS measurement process can then be written in the matrix form as:

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi^{-1} \mathbf{s} = \Theta \mathbf{s}, \quad (2)$$

where  $\mathbf{s} = \Psi \mathbf{x}$  is the sparse representation of the signal  $\mathbf{x}$  in some transformation domain  $\Psi$ , and  $\Theta = \Phi \Psi^{-1}$  is the sensing matrix. System in Equation (2) is underdetermined with a higher number of unknowns than the number of equations. Consequently, it has an infinite number of possible solutions. Basic assumption in the CS is that the observed signals have a sparse representation (i.e. only a small number of non-zero elements) in certain transformation domain. This assumption holds for most natural signals. However, solving Equation (2) to find the sparsest solution (i.e. the one with the smallest  $l_0$  norm) is NP-hard. Thus,  $l_1$ -norm solution is commonly used as a convex relaxation of the  $l_0$ -norm since it can be found using convex optimization algorithms [23].

CS reconstruction process can be modelled in its unconstrained form as:

$$\min_{\mathbf{s}} \|\mathbf{y} - \Phi \Psi^{-1} \mathbf{s}\|_2^2 + \lambda \|\mathbf{s}\|_1, \quad (3)$$



**Figure 3.** Flowchart of the proposed method. First, detailed calibration of the camera-projector system is performed, followed by the CS measurement process. Sparse optimization is applied to the CS measurements resulting in high-resolution sub-aperture reconstructions. Individual sub-aperture reconstructions are transformed from the projector coordinate system to the camera coordinate system using estimated perspective transformations. Finally, different post-capture effects can be applied to the reconstructed light field including refocusing, virtual aperture and viewpoint change.

where the left term corresponds to the Euclidean loss, while the right term corresponds to the  $l_1$ -regularization. In order for the CS reconstruction to be feasible, measurement matrix  $\Phi$  has to be incoherent with the transformation matrix  $\Psi$ , meaning that the rows of  $\Phi$  cannot sparsely represent the columns of  $\Psi$ , and vice versa. Any random measurement matrix, like a random Gaussian measurement matrix or a random binary matrix, is incoherent with any transformation matrix  $\Psi$  with high probability [16].

## 4. Experiments

In this section, we describe the CS measurement setup and a reconstruction algorithm that enables high resolution light field reconstruction from the CS measurements. Flowchart in Figure 3 shows the steps of the proposed CS light field reconstruction framework.

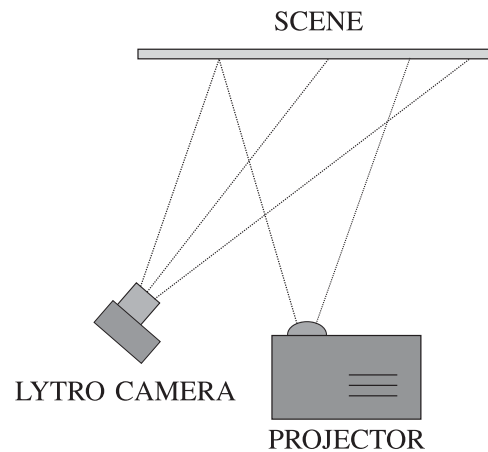
### 4.1. Measurement setup

The measurement setup proposed in this paper is a generalization of the compressive imaging setup proposed in [24]. It consists of a Lytro Illum light field camera and InFocus IN3118HD DLP projector, along with mechanical integration elements and a computer which is used for the CS reconstruction. All of the used components are unmodified off-the-shelf components. Measurement is performed in a darkroom to eliminate the influence of the ambient illumination. Camera and projector are remotely triggered in order to preserve the geometry of the measurement setup.

For simplicity, in our experiments we use a flat board with printed details as the scene of interest. We ensure that the scene has varying depth by carefully positioning the measurement setup. Camera is placed at an arbitrary angle (in our experiments at around  $45^\circ$ ) with respect to the scene, while the projector is kept perpendicular as depicted in Figure 4.

### 4.2. Camera-projector calibration

Calibration of the camera-projector system is a crucial step that precedes the CS measurement process. During



**Figure 4.** Visualization of the measurement setup. Camera is placed at an arbitrary angle with respect to the scene while the projector is kept perpendicular. Such setup guarantees varying depth in the scene from the camera viewpoint and also reduces the projector blur due to its limited depth of field.

the CS measurement process, projector is used to modulate the observed scene, while the camera captures the reflected intensities. Any mismatch between the computer generated measurement patterns and the patterns projected by the projector directly influences the reconstruction quality and is modelled as the multiplicative noise. The calibration process for the proposed measurement setup closely follows the procedure described in our previous work [24].

Therefore, camera focus, ISO and exposure values are manually set and kept fixed during the measurement process. Camera exposure time is set to a multiple of projector colour-wheel rotation period in order to overcome the aliasing effects of the colour-wheel. Frequency of the colour-wheel rotation for the IN3118HD projector is 120 Hz. Thus, we set the camera exposure time to  $1/60$  s. Projector manufacturers often reduce the required bandwidth in the projector design by using chroma subsampling. In chroma subsampling, the resolution of the colour channels is decimated, while the resolution of the luminance channel is preserved. Undesired intensity variations can occur when projecting patterns with resolution higher than

the resolution of the projector’s colour channels. Thus, we project measurement patterns that are  $2\times$  down-sampled to match the 4:2:2 chroma subsampling ratio of the IN3118HD. Since in our experiments only binary measurement patterns are used, gamma correction of the projector is not necessary.

CS assumes linearity of the measurement process. Consequently, we need to ensure that the camera sensor is working in linear mode. Lytro Illum camera stores raw sensor data as a 10-bit image in Lytro Raw Format (LFR) and these images are used in our experiments. Additionally, the camera-projector system introduces noise into the CS measurements. The overall noise in the measurement system consists of the camera noise (i.e. readout, dark current and photon noise [25]) and projector noise (i.e. stray projector light and background illumination). We perform overall noise estimation as proposed in [24] by projecting a sequence of eight ‘black’ projector images (i.e. zeros in R, G and B channels) and capturing them using the Lytro camera. Mean ‘black’ image  $I_b$  is calculated by averaging the captured images. Subtracting the mean ‘black’ image from each measurement image  $I_c$  cancels out the deterministic error pattern of the whole measurement system, resulting in the background corrected camera image  $\hat{I}_c$ :

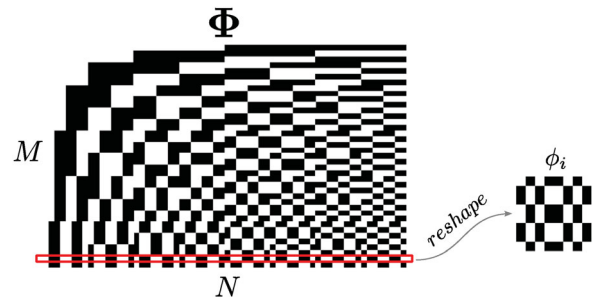
$$\hat{I}_c = I_c - I_b \quad (4)$$

### 4.3. Light field decoding

In order to decode the raw Lytro 2D lenslet image into the 4D light field representation, we closely follow the decoding procedure described in [26] and the Matlab implementation from [27]. Lytro camera has an unknown placement of the micro-lens array, with slight translational and rotational offsets in relation to the imaging sensor. Lytro cameras contain a database of white images which were taken through a diffuser. Due to the vignetting effect of each micro-lens (see Figure 2(c)), the micro-lens centres appear as the brightest spots in each micro-lens image. After detecting the micro-lens centroids, grid parameters are estimated by traversing the micro-lens centres using Delaunay triangulation.

Rotation of the micro-lens array is estimated by observing the angle of each individual row and column and by calculating the mean angle of the whole lenslet array. After the rotation is estimated, mean horizontal and vertical spacing between the micro-lens centroids is calculated. Raw sensor images are then re-sampled so that the micro-lens centres are aligned with the pixel centres. Additionally, re-sampling occurs during conversion of the hexagonal grid into an orthogonal grid.

The decoding procedure results in a light field with standard 4D parameterization. Light field obtained



**Figure 5.** Visualization of the Hadamard measurement matrix. Row vectors of the measurement matrix  $\Phi$  define the measurement patterns. Rows are rasterized into 2D images to be projected onto the scene.

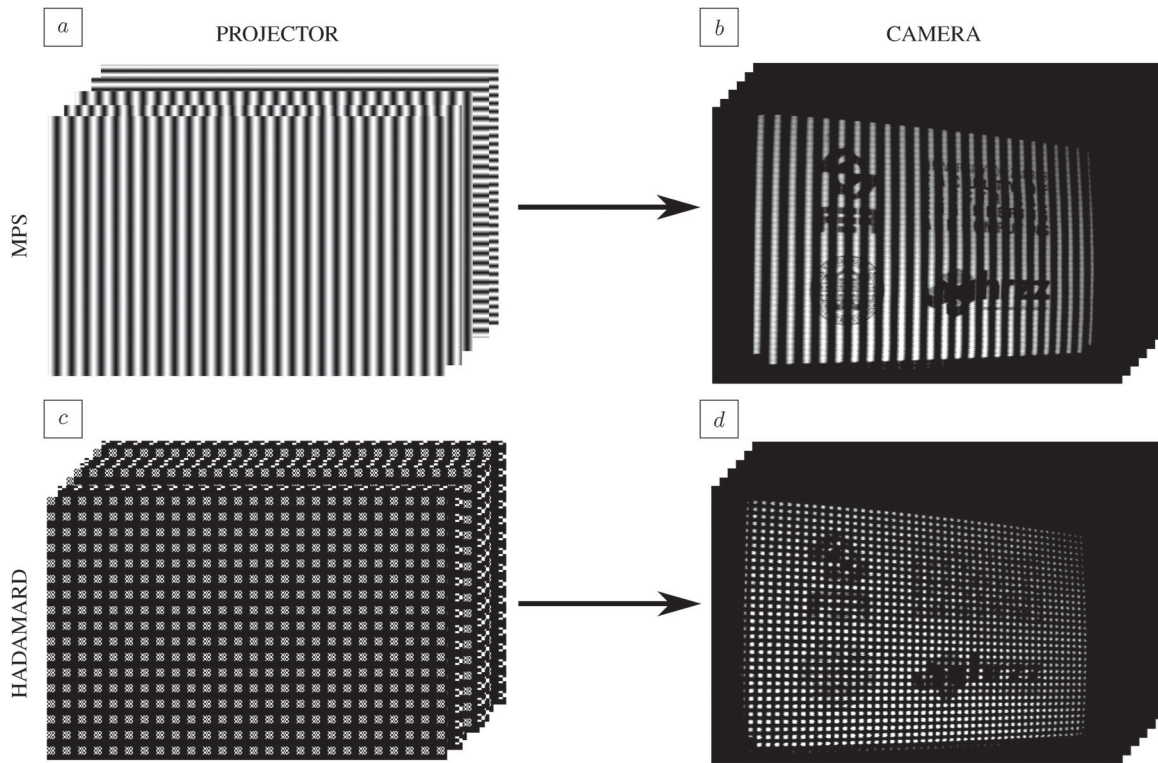
using the decoding procedure from [27] has angular resolution of  $15 \times 15$  and spatial resolution of  $626 \times 434$  pixels after the previously mentioned re-sampling steps.

### 4.4. Measurement process

In the experiments, we use Hadamard measurement matrix  $\Phi$  [28] shown in Figure 5, since it can be efficiently implemented in our measurement system using binary measurement patterns. As previously mentioned, this simplifies the calibration procedure by eliminating the need for the projector gamma correction. For the transformation matrix  $\Psi$ , we use discrete cosine transformation (DCT), as our scenes are approximately sparse in the DCT domain.

Number of CS measurements  $M$  and dimensionality of the observed signal  $N$  define the size of the measurement matrix  $\Phi$  and the size of the sparse optimization problem that needs to be solved. Row vectors of the measurement matrix  $\Phi$  define measurement patterns  $\phi_i$  to be projected onto the scene (see Figure 5). Pay attention that obtaining the CS measurements and performing sparse optimization for high-dimensional signals is time consuming and extremely computationally complex. To overcome this problem, we proposed phased CS measurement approach in [24]. The basic idea is to divide the original high-dimensional problem into a number of smaller sub-problems which can be measured and reconstructed independently. This approach significantly reduces the computational complexity and makes the CS reconstruction of high-dimensional signals feasible.

In our experiments, the projector image is partitioned into non-overlapping  $8 \times 8$  pixel blocks. In order to establish *per block* correspondences in the camera-projector system, we use multiple sine patterns which are shifted in phase (i.e. multiple phase shift, MPS approach [29,30]). MPS patterns are projected onto the scene independently from the CS measurement patterns and captured by the Lytro camera. Raw Lytro images of the MPS patterns are first

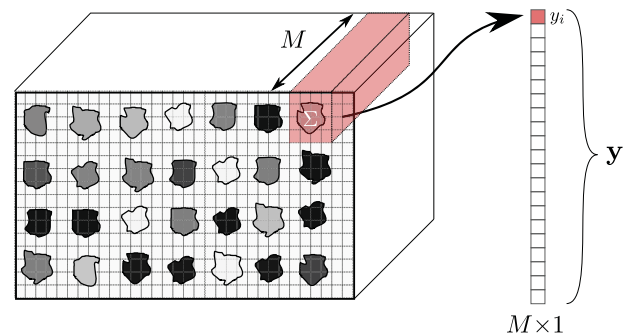


**Figure 6.** (a) A set of MPS patterns are projected onto the scene using projector. (b) Camera captures the MPS patterns from its viewpoint and camera-projector correspondences are calculated from the MPS images. (c) A set of Hadamard measurement patterns used in the CS measurement process. Note that in the visualization the projected patterns correspond to a single phase only. (d) Images of the scene modulated by the measurement patterns captured by the camera. CS reconstruction is applied on the acquired measurements (from all measurement phases) in order to obtain high-resolution scene reconstruction.

decoded into sub-aperture images and the correspondences are found for each sub-aperture image independently. Images corresponding to the projected MPS patterns, along with the images captured by the camera (for a single sub-aperture) are shown in Figure 6 (a ,b), respectively. After estimating and unwrapping the phase of the captured sine pattern, we find the exact correspondences between the projector and camera pixels by matching the corresponding phase values.

Next, we project Hadamard measurement patterns onto the scene and capture CS measurements using the Lytro camera as shown in Figure 6(c,d). In order to achieve the separability of the measurements, we project only a subset of blocks in a single measurement phase. In our case, the measurements are acquired by separating the blocks into 4 measurement phases.

A set of  $M$  measurement patterns (where  $M \ll N$ ) is projected onto the scene sequentially. Camera and projector lens introduce geometric spatial distortions, and the camera sensor captures the distorted projection of the scene. Captured projector blocks fall on several pixels of the camera sensor, since the camera and projector grids are arbitrary aligned. In order to isolate an individual measurement block, we use the previously obtained camera-projector correspondences. Intensities recorded by the camera sensor define the measurement results  $\mathbf{y}$ . Measurement results are obtained by summing the intensities of camera



**Figure 7.** CS measurement vector calculation for an individual block from a single measurement phase and for a single Lytro sub-aperture.  $M$  images of the observed scene modulated by the measurement patterns are captured. Measurement results are obtained by summing the intensities of camera pixels that correspond to a single projector block and by stacking them into a measurement vector  $\mathbf{y}$ . Similar procedure is performed for all measurement blocks and for all measurement phases.

pixels that correspond to the individual projected  $8 \times 8$  block. In Figure 7, calculation of the measurement vector for a single block in a single measurement phase is depicted. Previously described procedure is performed for each block of the projected measurement pattern.

#### 4.5. CS light field reconstruction

In our approach, CS reconstruction procedure is performed on the measurement results for each of the

Lytro sub-apertures. For the reconstruction, we use sparse optimization algorithm from SParse Modeling Software (SPAMS), an optimization toolbox for solving various sparse optimization problems [31]. Our sparse estimation problem is formulated following Equation (3). A single sub-aperture CS reconstruction results in a high-resolution image of the observed scene from the projector viewpoint.

In order to acquire the high resolution reconstruction from the camera viewpoint, we use homography transformation.

In the following, we give a short overview of the homography estimation based on [32]. Homography is estimated between each of the Lytro camera sub-apertures and the projector using the previously obtained correspondences. Under homography, we can write the transformation of the corresponding points in 3D from the camera to projector coordinate system as:

$$\mathbf{X}_2 = \mathbf{H}\mathbf{X}_1, \quad \mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}. \quad (5)$$

In the image planes, using homogeneous coordinates, we can write:

$$\lambda_2 \mathbf{x}_2 = \mathbf{H} \lambda_1 \mathbf{x}_1. \quad (6)$$

Equation (6) denotes direct mapping between the points in the camera and projector planes up to a scale. To estimate the unknown homography  $\mathbf{H}$ , we start from  $\mathbf{x}_2 \sim \mathbf{H}\mathbf{x}_1$ . Using homogeneous coordinates we can write the following matrix equation:

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \Leftrightarrow \mathbf{x}_2 = \mathbf{H}\mathbf{x}_1. \quad (7)$$

After rewriting Equation (7) in inhomogeneous coordinates with  $x'_2 = x_2/z_2$  and  $y'_2 = y_2/z_2$  and by setting  $z_1 = 1$ , we get the following system:

$$x'_2(H_{31}x_1 + H_{32}y_1 + H_{33}) = H_{11}x_1 + H_{12}y_1 + H_{13}, \quad (8)$$

$$y'_2(H_{31}x_1 + H_{32}y_1 + H_{33}) = H_{21}x_1 + H_{22}y_1 + H_{23}. \quad (9)$$

Given a set of corresponding points between the camera and the projector, we can form the following linear system of equations:

$$\mathbf{A}\mathbf{h} = \mathbf{0}, \quad (10)$$

where

$$\mathbf{A} = \left( \mathbf{a}_{x_1}^T \mathbf{a}_{y_1}^T \dots \mathbf{a}_{x_N}^T \mathbf{a}_{y_N}^T \right)^T \quad (11)$$

$$\mathbf{h} = (H_{11}, H_{12}, H_{13}, H_{21}, H_{22}, H_{23}, H_{31}, H_{32}, H_{33})^T \quad (12)$$

$$\mathbf{a}_x = (-x_1, -y_1, -1, 0, 0, 0, x'_2 x_1, x'_2 y_1, x'_2)^T \quad (13)$$

$$\mathbf{a}_y = (0, 0, 0, -x_1, -y_1, -1, y'_2 x_1, y'_2 y_1, y'_2)^T. \quad (14)$$

Equation (10) can be solved using homogeneous linear least squares method in order to obtain the estimate of the homography matrix. In our experiments, we use MATLAB function *fitgeotrans* to calculate the homography matrix from the estimated phase values in camera images, and the known projected phase values.

After the homography matrix is obtained, we are able to transform the high-resolution image reconstruction from the projector viewpoint to the camera viewpoint. Same procedure is repeated for each of the Lytro camera sub-apertures. This results in a high-resolution reconstruction of the light field. Finally, individually reconstructed sub-aperture images are then reordered into the 4D light field format. The presented approach thus preserves all post-capture capabilities of the Lytro camera which will be demonstrated in the next section.

## 5. Results

### 5.1. Improved spatial resolution of light field

In this section, we discuss the results obtained using the proposed framework. First, we show the improvement in the spatial resolution of the captured light field. We select the central sub-aperture from the Lytro camera for visualization purposes, but similar observations hold for other sub-apertures and we instruct the interested reader to check the supplementary materials for additional visualizations. In Figure 8(a), a single measurement from the Lytro camera sensor is shown. Pay attention that this visualization corresponds to the per block measurement results. Figures 8(b–d) show the CS reconstructions using different number of measurements, i.e.  $M = 24, 32, 64$ . As noted before,  $M$  corresponds to the total number of measurements obtained using the proposed setup. The reconstruction obtained using 64 measurements corresponds to the Nyquist sampling rate.

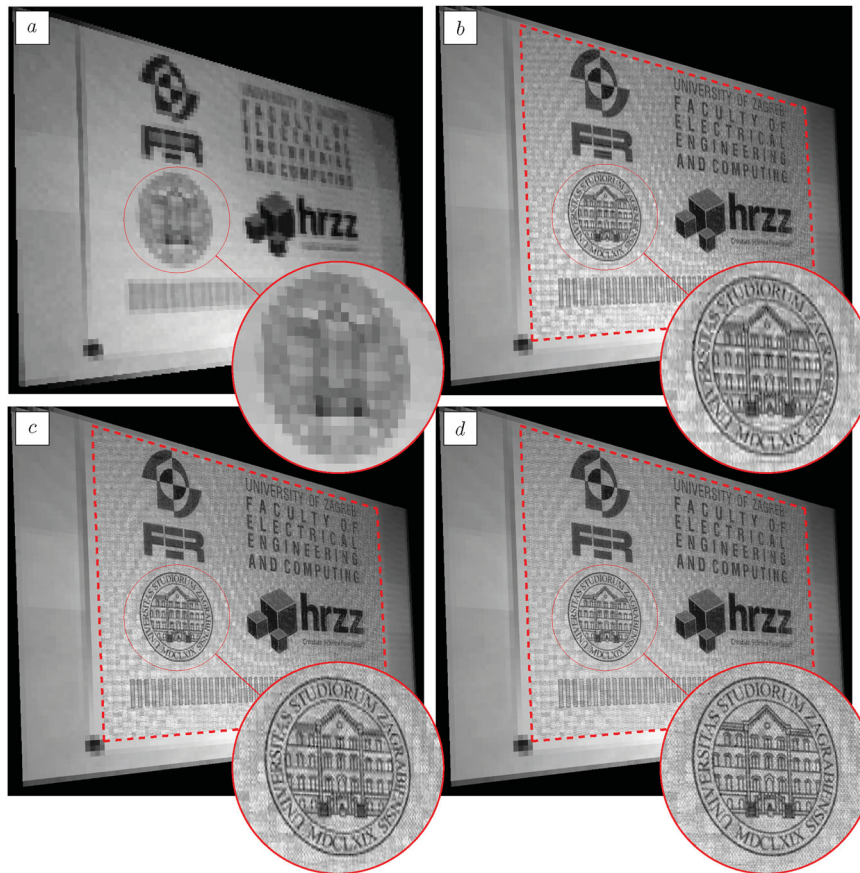
### 5.2. Viewpoint change

After we have successfully increased the spatial resolution of each of the Lytro sub-apertures, we focus on the post-capture manipulations of the reconstructed light field. One possible post-capture manipulation in light field imaging is change of the viewpoint. Sub-aperture images are formed by extracting the same pixel under each micro-lens.

By traversing the sub-aperture images and selecting a single sub-aperture at a time, we get the scene reconstruction from different viewpoints. In Figure 9, a visualization of the central sub-aperture along with the tiled light field format is shown. Each tile corresponds to a single high-resolution sub-aperture image.

In Figure 10, a composite of two sub-aperture images is shown. Pay attention that only an enlarged part of the reconstructed scene is shown. Since significant





**Figure 8.** Increasing spatial resolution of light field using compressive sensing. (a) Visualization of a single measurement result in the CS measurement process. CS reconstructions of the central sub-aperture from sub-sampled number of measurements are shown in the sub-figures (b)–(d) where the number of measurements are  $M = \{24, 32, 64\}$  respectively. CS reconstructed insets are outlined by the dashed line in the presented images.



**Figure 9.** Reconstructed image of the central Lytro sub-aperture along with the tiled visualization of the reconstructed light field shown in the inset. In order to obtain the tiled light field, all the reconstructed sub-apertures are arranged into the  $15 \times 15$  grid, where each tile corresponds to a single sub-aperture.

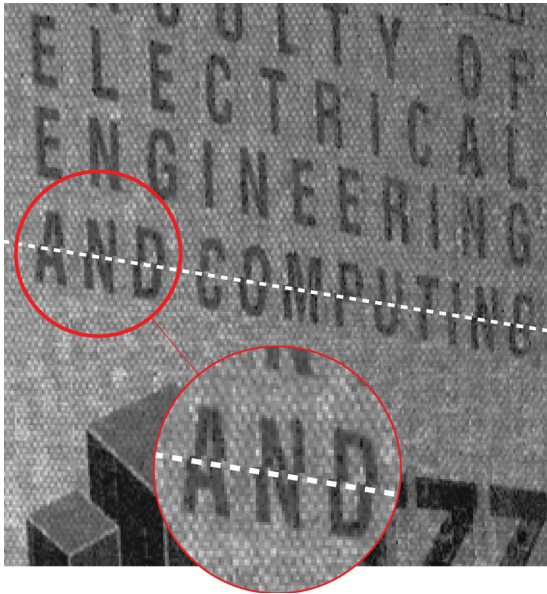
vignetting occurs at the boundary sub-apertures, for the visualization purposes, sub-apertures with indices (8, 3) and (8, 13) are selected. This corresponds to the distant sub-apertures in the middle row of the  $15 \times 15$

grid. Dashed line denotes the boundary between the two images. Notice the slight offset of the letters near the boundary in the magnified detail. This demonstrates the viewpoint change in the reconstructed light field. Better visualization of the viewpoint change is available in GIF format in the supplementary material.

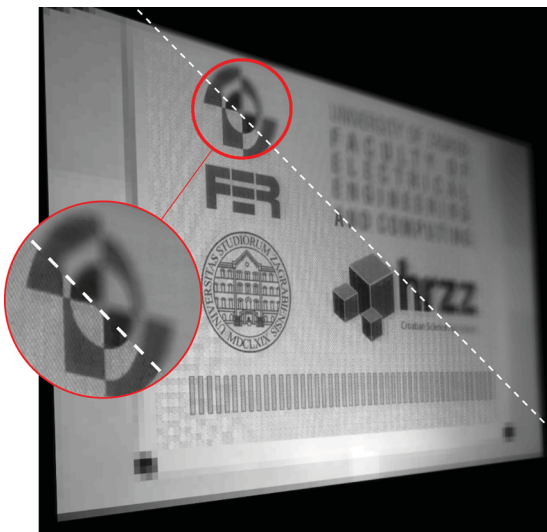
### 5.3. Post-capture refocusing and virtual aperture

Refocusing is another potential post-capture manipulation of the light field. There are several algorithms for refocusing light fields [8,33], and in this paper we use the shift-and-sum algorithm from [34,35]. Refocusing using shift-and-sum algorithm is done by numerically approximating the integral from Equation (1), and conceptually is a summation of dilated and shifted versions of the sub-aperture images over the entire  $(u, v)$  aperture.

Virtual change of the camera aperture is performed in a similar manner. By summing different number of sub-apertures without shifting them, we can obtain an arbitrary aperture. By summing all of the sub-apertures, we get the original physical aperture of the Lytro camera. In Figure 11, a composite of two images is shown. In the first image, the focus plane is located at the board plane resulting in a sharp reconstruction of the scene.



**Figure 10.** Visualization of the horizontal offset between the Lytro sub-aperture reconstructions at positions (8,3) and (8,13). Since boundary sub-apertures have significant vignetting, we do not visualize the leftmost and rightmost sub-aperture, but the near-boundary sub-aperture. Visualization is a composite of two sub-apertures where the dashed line denotes the boundary between the views. Notice the slight horizontal offset visible in the enlarged detail.



**Figure 11.** Visualization of the post-capture light field refocusing. A composite of two images is shown, where one image is focused on the scene, while the other image is focused further away from the scene. Dashed line denotes the boundary between the two images. Better visualization can be found in the supplementary material in animated *GIF* format.

In the second image, the focus is set further away from the scene resulting in blurry reconstruction. Dashed line denotes the boundary between the images. In the magnified detail, the difference between the two images is clearly visible. Better visualization is available in the supplementary material as a live *GIF* image.

## 6. Conclusion

In this paper, we proposed a novel compressive sensing framework for high-resolution light field reconstruction. The original 40MP sensor resolution of the Lytro Illum light field camera is split between  $15^2 = 225$  angles, thus resulting in rather small spatial resolution after decoding the raw Lytro sensor image into the 4D light field format.

In order to overcome the spatio-angular trade-off of the light field acquisition, we designed an off-the-shelf measurement setup consisting of a light field camera and a digital projector. After calibration, a set of random patterns was projected on a static scene. Compressive sensing approach was used for a successful reconstruction of 225 Lytro sub-apertures in the projector resolution. The potential resolution of the virtual sensor in the proposed measurement system is up to 445MP under assumption of a full HD projector ( $1920 \times 1080$ ). Post-capture capabilities of the light field, including view-point change, refocusing and virtual aperture, are preserved using the proposed method.

Since the correspondences between the projector and the light field camera views were modelled as homography, the scene is limited to a plane of an arbitrary angle in 3D space. Our results in high-resolution light field reconstruction have practical significance, but are still limited to a rather simple planar scene. Hence, this paper presents a proof-of-concept of an original approach that will be generalized to an arbitrary scene in our future research.

## Acknowledgments

The authors would like to thank Matea Donlic for the help and fruitful discussions during this research.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported in part by the Hrvatska zaklada za znanost (Croatian Science Foundation) under Projects IP-2014-09-2625, IP-2019-04-6703, and in part by European Regional Development Fund KK.01.1.1.01.0009 (DAT-ACROSS).

## ORCID

Ivan Ralašić  <http://orcid.org/0000-0003-4479-9099>

Damir Seršić  <http://orcid.org/0000-0001-5008-5047>

## References

- [1] Adelson EH, Bergen JR. The plenoptic function and the elements of early vision. In: Landy MS, Movshon AJ, editors. Computational models of visual processing. Cambridge (MA): MIT Press; 1991. p. 3–20.
- [2] Lippman G. La photographie integrale. Comptes-Rendus. Academie des Sciences; 1908. p. 446–551.

- [3] Levoy M, Hanrahan P. Light field rendering. Proceedings of the 23rd annual conference on Computer graphics and interactive techniques – SIGGRAPH '96; ACM Press; 1996.
- [4] Liang C-K, Lin T-H, Wong B-Y, et al. Programmable aperture photography: multiplexed light field acquisition. *ACM Trans Graph (TOG)*. 2008;27:55.
- [5] Marwah K, Wetzstein G, Bando Y, et al. Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM Trans Graph (Proc SIGGRAPH)*. 2013;32(4):1–11.
- [6] Wilburn B, Joshi N, Vaish V, et al. High performance imaging using large camera arrays. *ACM Trans Graph (TOG)*. 2005;24:765–776.
- [7] Ng R, Levoy M, Brédif M, et al. Light field photography with a hand-held plenoptic camera. *Comput Sci Tech Report CSTR*. 2005;2(11):1–11.
- [8] Ng R. Digital light field photography [PhD thesis]. Stanford: Stanford University; 2006.
- [9] Wang X, Li L, Hou G. High-resolution light field reconstruction using a hybrid imaging system. *Appl Opt*. Mar. 2016;55:2580.
- [10] Alam MZ, Gunturk BK. Hybrid light field imaging for improved spatial resolution and depth range. *Mach Vis Appl*. Jan 2018;29:11–22.
- [11] Yoon Y, Jeon H-G, Yoo D, et al. Learning a deep convolutional network for light-field image super-resolution. 2015 IEEE International Conference on Computer Vision Workshop (ICCVW); IEEE; Dec. 2015.
- [12] Gul MSK, Gunturk BK. Spatial and angular resolution enhancement of light fields using convolutional neural networks. *IEEE Trans Image Process*. May 2018;27:2146–2159.
- [13] Zhang S, Lin Y, Sheng H. Residual networks for light field image super-resolution. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2019. p. 11046–11055.
- [14] Shan X. 4D light field reconstruction and scene depth estimation from plenoptic camera raw images [PhD thesis]. Galway: School of Physics, Science Faculty, National University of Ireland; 2016.
- [15] Candes EJ, Tao T. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans Inform Theory*. Dec 2006;52:5406–5425.
- [16] Baraniuk RG. Compressive sensing [lecture notes]. *IEEE Signal Process Mag*. July 2007;24:118–121.
- [17] Rish I, Grabarnik G. Sparse modeling: theory, algorithms, and applications. Boca Raton (FL, USA): CRC Press; 2014.
- [18] Rani M, Dhok S, Deshmukh R. A systematic review of compressive sensing: concepts, implementations and applications. *IEEE Access*. 2018;6:4875–4894.
- [19] Duarte MF, Davenport MA, Takhar D, et al. Single-pixel imaging via compressive sampling. *IEEE Signal Process Mag*. 2008;25(2):83–91.
- [20] Kulkarni K, Lohit S, Turaga P, et al. ReconNet: non-iterative reconstruction of images from compressively sensed measurements. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); IEEE; Jun 2016. p. 449–458.
- [21] Ralašić I, Tafro A, Seršić D. Statistical compressive sensing for efficient signal reconstruction and classification. 2018 4th International Conference on Frontiers of Signal Processing (ICFSP). Sep. 2018. p. 44–49.
- [22] Ralašić I, Seršić D. Real-time motion detection in extremely subsampled compressive sensing video. 2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA) (IEEE ICSIPA 2019); Kuala Lumpur, Malaysia: Sept. 2019.
- [23] Donoho DL. For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution. *Commun Pure Appl Math*. 2006;59(6):797–829.
- [24] Ralašić I, Seršić D, Petrinović D. Off-the-shelf measurement setup for compressive imaging. *IEEE Trans Instrum Meas*. 2018;68(2):502–511.
- [25] Seršić D, Sović A, Menoni CS. Restoration of soft x-ray laser images of nanostructures. *Opt Express*. May 2014;22:13846.
- [26] Dansereau DG, Pizarro O, Williams SB. Decoding, calibration and rectification for lenselet-based plenoptic cameras. 2013 IEEE Conference on Computer Vision and Pattern Recognition; IEEE. Jun 2013. p. 1027–1034.
- [27] Dansereau DG. Light field toolbox for Matlab. 123, 2 2015. [cited 2019, 05]. Version v0.4.
- [28] Evangelaras H, Koukouvinos C, Seberry J. Applications of hadamard matrices. *J Telecommun Inform Technol*. 2003;2:3–10.
- [29] Zuo C, Feng S, Huang L, et al. Phase shifting algorithms for fringe projection profilometry: a review. *Opt Lasers Eng*. 2018;109:23–59.
- [30] Petkovic T, Pribanic T, Donlic M. Temporal phase unwrapping using orthographic projection. *Opt Lasers Eng*. 2017;90:34–47.
- [31] SPAMS Sparse Modeling Software. [cited 2019, 04]. Available from: <http://spams-devel.gforge.inria.fr/>
- [32] Hartley R, Zisserman A. Multiple view geometry in computer vision. NY, USA: Cambridge University Press; 2003.
- [33] Dansereau DG, Pizarro O, Williams SB. Linear volumetric focus for light field cameras. *ACM Trans Graph*. 20, Mar. 2015;34:15:1–15:15.
- [34] Vaish V, Wilburn B, Joshi N, et al. Using plane+ parallax for calibrating dense camera arrays. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004; Vol. 1, IEEE; 2004. pp. I–I.
- [35] Levoy M, Chen B, Vaish V, et al. Synthetic aperture confocal imaging. *ACM Transactions on Graphics (ToG)*. Vol. 23, ACM; 2004. p. 825–834.