# Plenopticカメラにおけるエンド・トゥ・エンド システムモデル

## End-to-end System Model for Plenoptic Cameras

キャサリン バークナー* Kathrin BERKNER
リンフェイ メング* Lingfei MENG
サプーナ シュオッフ* Sapna A. SHROFF
イバナ トシック* Ivana TOSIC

**要 旨**

Plenopticカメラのデザインスペースのためのシステムモデルを提案する．本モデルには，先端的な画像再構成手法と応用に応じた性能メトリクスに加え，光学および検出器サブシステムに関する詳細情報が含まれている．このモデルにより，物体から発せられた光からカメラセンサへの線形のマッピングを表現する，システム伝達マトリクスが求まる．画像形成に関するこの線形モデルに基づき，線形逆問題を解く理論の概念を活用することで，空間情報および分光情報の再構成法を提案する．さらに，異なるデータ再構成法の評価のための本システムモデルの利用法を示す．

**ABSTRACT**

We introduce a system model for the design space of a plenoptic camera, including detailed information about the optical and detector subsystems, as well as advanced image reconstruction techniques and application-specific performance metrics. A system transfer matrix is derived that describes a linear mapping of light originating from an object onto the camera sensor. Based on that linear model for the forward image formation, reconstruction method for spatial and spectral information are proposed utilizing concepts from the theory of solving linear inverse problems. The use of the system model for evaluation of different data reconstruction methods is demonstrated.

* リコー イノベーションズ コーポレーション
Ricoh Innovations Corporation

## 1. Introduction

A recently developed computational imaging system is a plenoptic camera, which provides additional functionalities compared to a standard camera, such as instantaneous multi-spectral imaging, refocusing, and 3D imaging[1-3]. Those functionalities are achieved via insertion of a microlens array close to the detector, an optional spectral filter array inserted in the main lens, and use of advanced image processing algorithms (Fig.1). Plenoptic sensor data contain light field information of a scene, multiplexed by the optical system and the detector. The trade-offs between spatial, depth, and spectral resolution are determined by the characteristics of the optical system and the detector, as well as the imaging application at hand. In the literature, those trade-offs have been studied using first-order geometric models for the optical system, mostly concentrating on either the optical[1-3] or the signal processing system components[4,5].

When facing the problem of designing a plenoptic camera system, performance needs to be predicted in order to make appropriate design choices. In our experience, performance, however, is difficult to predict since the dimensionality of the design space is large, and related design trade-offs not very clearly analyzed by the published models. In order to characterize the design space of a plenoptic camera system, we built an end-to-end system model and implemented it in a simulation environment. In that environment, the design space and its tradeoffs can be explored and characterized using application-specific system metrics. In the following we will describe the system model as well as details regarding the adaptation of processing components to specific system metrics.

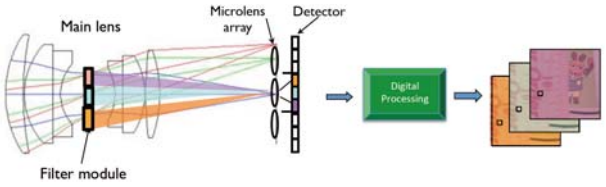


Fig.1 Overview of plenoptic camera with spectral filters inserted into main lens.

## 2. Understanding capture of spectral and spatial scene information by a plenoptic camera sensor.

### 2-1 Overview of plenoptic system model

During a camera design phase, the question arises of how accurate the measurements of the light modalities obtained with a plenoptic camera will be, what effect certain system parameters have on the overall system performance, and what algorithms to apply to sensor data. To answer these questions, we built an end-to-end system model including models for source, optics, detector, digital processing and application-specific performance metrics (Fig.2). The optics model includes a lens model for the main lens as well as for the microlens array, which consists of many small lenslets with diameters typically smaller than 250 microns. Using our system model, we are able to evaluate the spectral and spatial performance of a plenoptic camera through simulation of sensor data and design of system-specific reconstruction algorithms. Spatial performance evaluates resolution of small details in the image, whereas spectral performance evaluates accuracy of wavelength-specific information in the scene.

For a multispectral plenoptic camera, wavelength-dependent effects such as chromatic aberration can be analyzed. Chromatic aberration introduced by the main optics and the lenslets can lead to spectral multiplexing of wavelength information at the detector.

We also model the multiplexing of spatial information on the sensor in a small area under a single lenslet using wave propagation techniques. The simulated sensor data allow for development of reconstruction algorithms to recover the initial spectral and spatial scene information.

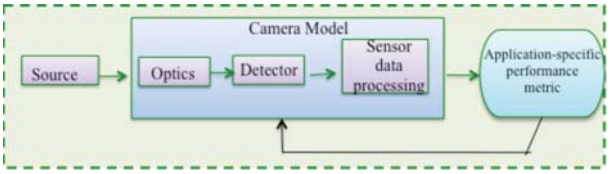


Fig.2 Overview of camera system model.

## 2-2 Spectral information model

Chromatic aberration in the optical system and pixelation at the sensor causes spectral multiplexing in the captured sensor data. This spectral multiplexing can be described by a linear model including a transfer matrix that contains detailed information of the multiplexing of spectral components when passing through the optical system. This transfer matrix depends on the optical characteristics of the main lens, the microlens characteristics, and sensor parameters[6]. With knowledge of the transfer matrix, we can define the requirements for signal processing technologies in the model to be able to solve a linear inverse problem. In 6), we demonstrated that using the pseudo inverse of the transfer matrix we could improve the spectral reconstruction compared to simple remapping algorithms.

In the source model the radiance reflected from the object is calculated based on irradiance of light and reflectance of object. Mean and covariance of the radiance are evaluated and propagated to a camera model.

A camera model is derived based on optical response of the system and a detector model. A geometric approximation of the optical response is obtained via ray tracing. The radiance originating from an object point passed through optics is converted to digital output measurements for all the sensor pixels in a super-pixel behind a lenslet

$$\mathbf{x} = \mathbf{F}\mathbf{b} + \mathbf{N}_{photon}(\mathbf{b}) + \mathbf{N}_{system}$$

where **x** is a vector containing sensor data in each super-pixel, **F** is a system response matrix, **b** is a vector containing the spectral intensity values, $\mathbf{N}_{\text{photon}}$ is the signal-dependent shot noise, and $\mathbf{N}_{\text{system}}$ is the signal independent noise, such as read noise and quantization noise. Embedding of statistical system properties and detailed optical characteristic into the plenoptic camera model is a major improvement over the simplistic models introduced in 3) and 7).

Based on our spectral multiplexing model, we can design algorithms to reconstruct the original spectral information of an object location. A spectral feature vector **b*** is reconstructed from the plenoptic sensor data vector **x** via linear transformation

$$\mathbf{b}^* = \mathbf{\Psi}\mathbf{x},$$

where $\mathbf{\Psi}$ is a spectral reconstruction matrix. In 6) we proposed a system-dependent spectral demultiplexing algorithm. The output signal is demultiplexed based on a estimated system response matrix $\hat{\mathbf{F}}$. This matrix is determined from calibration experiments, where narrow-band light is passed through the camera and its sensor response captured[8]. The spectral features are reconstructed by taking a pseudoinverse of $\hat{\mathbf{F}}$, i.e.,

$$\mathbf{\Psi} = (\hat{\mathbf{F}}^T\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}^T.$$

The performance can be evaluated in terms of spectral reconstruction error, signal-to-noise ratio, or by classification accuracy metric.

## 2-3 Spatial information model

### 2-3-1 Image formation

Object and sensor space in a plenoptic system are related by the system point response function called the pupil image function (PIF)[9], which is analogous to the point spread function (PSF) in conventional cameras.

This PIF, which includes optical properties of the main lens and the microlenses, is a highly spatially varying function with significant variations in a superpixel area, compared to a conventional PSF with slow variation across the entire field of view. Given an optical system and an object plane position, its PIF can be obtained using the principles of Fourier optics[9]. The collection of all the PIF responses for different points in the object space comprises the system matrix **A**. Fig.3 shows example of simulated PIF responses on the sensor for object points close to the optical axis. The images are obtained by propagating light originating from an object point through the main and the on-axis lenslet onto the sensor. The left image shows the PIF response for an on-axis object point, the middle and right image show PIF responses for points located a small distance away from the optical axis, demonstrating the highly spatially varying characteristics of the PIF response under a single lenset. This characteristic cannot be explored with the geometric optics models from 1), 2) and 5).

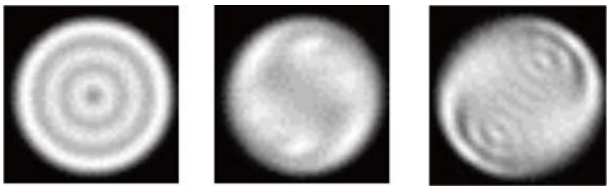

Fig.3 Examples of optical responses captured at the sensor for object points imaged through the on-axis lenslet.

If we denote the image at the sensor as **y** (in a vectorized form) and the object points as **x**, we can formulate the image acquisition process of our linear plenoptic system as:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \eta,$$

where **η** represents system noise[9]. The reconstruction problem is thus to find **x**, given **y** and **A**. The PIF matrix, however, can be rank deficient and recovering a high resolution object data represents a difficult inverse problem. This is particularly the case when the object is in focus at the lenslet array, as in this case there is no parallax between light rays and superresolution methods such as 5) are not applicable. The focused case is of particular interest for multi-spectral imaging systems.

### 2-3-2 Image reconstruction

To solve an under-determined linear system we can use least-squares-based solutions or more advanced optimization techniques. The least-squares-solutions provide good reconstruction in the absence of noise, but degrade when sensor noise is present[9]. One way to enhance the robustness to noise is to incorporate some prior information about the signal or image, such as sparsity, i.e. the embedding of low-dimensional structure in high-dimensional signals. The recent developments of theory and algorithms for processing of sparse signals has lead to the discovery of computational tools to recover low-dimensional structures in high-dimensional data. The tools often use complex optimization methods to solve ill-conditioned linear inverse problems given certain priors on data sparsity[10]. Such prior assumes that signal is sparse in a certain dictionary $\mathbf{\Phi}$, i.e., that in the signal model $\mathbf{x} = \mathbf{\Phi}\mathbf{c}$, the vector of coefficients **c** has a small number of non-zero entries. Compressive Sensing (CS) theory addresses the problem of the reconstruction of sparse signals from linear measurements[10]. Following the introduced notation, the CS reconstruction problem is to find a sparse estimate for **c** from measurements **y** such that $\mathbf{y} = \mathbf{A}\mathbf{\Phi}\mathbf{c} + \eta$.

When certain conditions on **A** and $\mathbf{\Phi}$ are met, it has been shown that sparse **c** can be reconstructed by solving the following convex optimization problem:

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \left[ \|\mathbf{y} - \mathbf{A}\mathbf{\Phi}\mathbf{c}\|_2^2 + \lambda \|\mathbf{c}\|_1 \right],$$

where λ is a trade-off parameter between the level of sparsity and the fidelity of signal reconstruction[10]. This optimization problem can be solved efficiently using interior point or gradient methods. However, CS theory requires that the coherence between **A** and $\mathbf{\Phi}$ is small.

The mutual coherence between **A** and $\mathbf{\Phi}$ is defined as the following equation[11]:

$$\mu(\mathbf{A}, \mathbf{\Phi}) = \max_{i,j} |\langle \mathbf{a}_i, \phi_j \rangle| << 1$$

where $\mathbf{a}_i$ is the i-th row of **A** , $\varphi_j$ is the j-th column of $\mathbf{\Phi}$.

If the signals are not directly observed, but are measured through a measurement matrix **A**, estimation of **c** requires that matrices **A** and $\mathbf{\Phi}$ have small mutual coherence. This condition influences the dictionary learning process because we need to find a dictionary that not only well describes the data, but is also incoherent with the system measurement matrix **A**.

In 12), we propose a new algorithm that achieves such learning. It is a two-step algorithm, which alternates between estimating sparse coefficient vectors and optimizing the dictionary. The algorithm is summarized in Fig.4. The sparse coefficient vector **c** for each column of the training data matrix **X** is stored as a column in the dictionary coefficient matrix **C**. Gradient methods are used to solve the two optimization problems for **C** and **Φ**.

**Algorithm 1** Dictionary learning for incoherent sampling

Input: training data $\mathbf{X}_t$, measurement matrix $\mathbf{A}$, parameters $\sigma, \lambda, \delta, p, L$ (dictionary size), $B > 4L$
$[N, Q] = \text{size}(\mathbf{X}_t)$; $M = \text{size}(\mathbf{A}, 1)$
Initialize dictionary at random: $\mathbf{\Phi} \sim \mathcal{U}^{N \times L}(-0.5, 0.5)$
Run learning for $p$ iterations (or until convergence):
**for** $i = 1 \rightarrow p$ **do**
  Randomly select $B$ training signals: $\mathbf{X} = \mathbf{X}_t(:, s), s = [t], t \sim \mathcal{U}^{B \times 1}(0, Q)$
  Generate noisy measurements: $\mathbf{Y} = \mathbf{A}\mathbf{X} + \boldsymbol{\eta}$, $\boldsymbol{\eta} \sim \mathcal{N}^{M \times N}(0, \sigma)$
  Initialize coefficients: $\mathbf{C}_0 = \mathbf{0}$
  Solve: $\hat{\mathbf{C}} = \arg\min_{\mathbf{C}} \left[ \|\mathbf{Y} - \mathbf{A}\mathbf{\Phi}\mathbf{C}\|_2^2 + \lambda \|\mathbf{C}\|_1 \right]$
  Solve: $\hat{\mathbf{\Phi}} = \arg\min_{\mathbf{\Phi}} \left[ \frac{1}{B} \|\mathbf{X} - \mathbf{\Phi}\hat{\mathbf{C}}\|_F^2 + \delta \|\mathbf{A}\mathbf{\Phi}\|_F^2 \right]$
  Normalize columns of $\hat{\mathbf{\Phi}}$ : $\hat{\phi}_j := \frac{\hat{\phi}_j}{\|\hat{\phi}_j\|_2}, \forall j \in [1, L]$
  $\mathbf{\Phi} := \hat{\mathbf{\Phi}}$
**end for**
Output: $\mathbf{\Phi}$

Fig.4 Dictionary learning algorithm for plenoptic data representation and reconstruction from 13).

## 3. Reconstruction results using the system model

### 3-1 Spectral information extraction

The derived model, including spectral demultiplexing, is used to evaluate the system performance for a prototype consisting of a commercially available DSLR camera and a filter array with four narrow-band spectral filters with center wavelengths 450, 540, 570, and 650nm (Fig.5). Spectral calibration is then performed to construct the system response matrix. To calibrate the system response, monochrome light corresponding to the wavelength of each spectral filter is sent through the main lens. The response of each filter can then be captured and used to form the system response matrix[14]. Examples of the response captured by sending monochrome light with central wavelength of 540 nm and 650 nm are shown in Fig.6.

Our spectral demultiplexing method is compared with the single pixel extraction method[7], in which a single pixel in the image of each filter cell on the sensor with maximum intensity is selected. Reflectance values measured from different patches in a color checker are used as ground truth data. The reconstructed spectral intensity values are normalized to reflectance based on a reference image captured on a Labsphere reflectance target. The spectral reconstruction error is computed at four different wavelengths, and by averaging the values calculated based on the white, black, red, orange, green, and blue patches. In this evaluation the reflectance is reconstructed from only one lenslet. The results based on single pixel approach and demultiplexing are compared in Figs.7 and 8. It is seen that our demultiplexing approach is comparable to the single pixel extraction approach, showing, that we can compensate for the chromatic system distortions by demultiplexing the sensor data according to the lens characteristics. The

system performance is further evaluated based on the SNR. The SNR is computed based on the image of a white object captured with two different exposure settings. The results are shown in Fig.7. It is shown that the spectral demultiplexing method reconstruct the spectral information with much higher SNR.

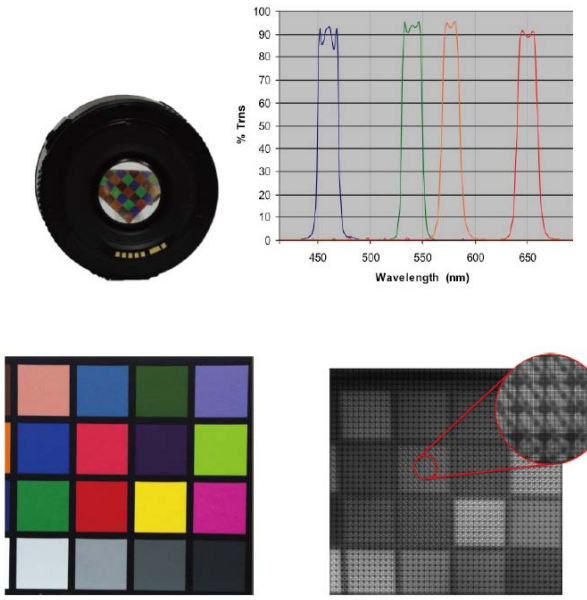


Fig.5 Prototype filter array in main lens with the corresponding spectral filter responses (top), color checker target and sensor data captured spectrally coded plenoptic camera (bottom).

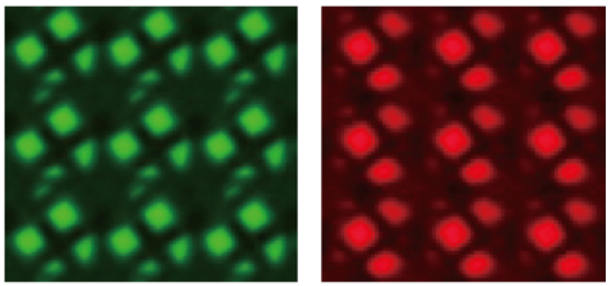


Fig.6 Plenoptic calibration data for two different wavelengths: a) 540nm, b) 650nm.

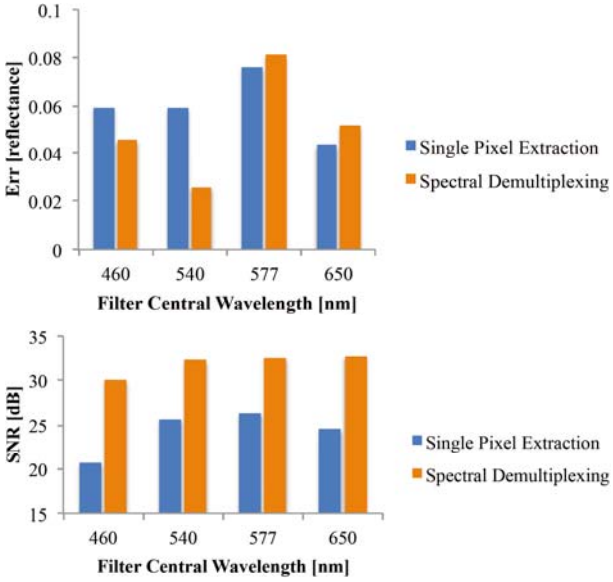


Fig.7 Performance evaluation of spectral plenoptic camera, measuring spectral reconstruction error (top), and SNR (bottom).

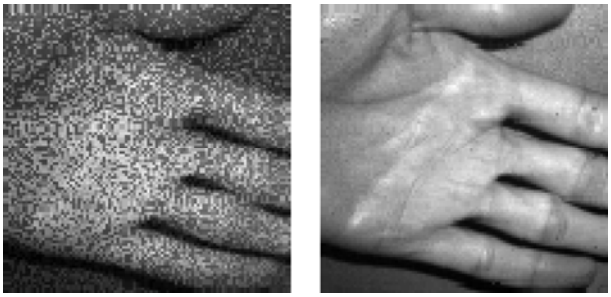

Fig.8 Spectral reconstruction of an image of a hand capture at 650nm: single pixel method (left), proposed demultiplexing method (right).

## 3-2 Spatial information extraction

The use of the system for the recovery of spatial information from plenoptic sensor data is demonstrated using software simulation. We have first simulated the PIF matrix for one on-axis lenslet using the wave-propagation analysis of the non-aberrated plenoptic system[9]. The resulting linear system simulates image formation for the case of a planar object that is in focus at the microlens array plane. For dictionary learning we use a training set **X** of video frames from a natural movie database (as used previously in 11)). We have learned a

dictionary of atoms of size 40x40 with L = 1600 atoms. This block size is chosen such that a block is imaged by the main lens on exactly one lenslet covering 52x52 sensor pixels. In each iteration we have selected a batch of 6400 blocks of size 40x40. Each block has been reshaped into a vector and placed into a column of **X**. We have then simulated the plenoptic imaging process as $\mathbf{y} = \mathbf{A}\Phi\mathbf{c} + \eta$, where **A** is the PIF matrix and **η** is white Gaussian noise of SNR= 60dB. An example PIF matrix is shown in the top left of Fig.9. Two example images contained in the training set are shown in the top right of Fig.9. The vectorized form of those intensity data form the columns of the matrix **X**. The bottom pictures in Fig.9 show the result of the learning algorithm, based on the optimized coefficient dictionary **Φ**: three sensor space elements, i.e. three column data of **AΦ** (left), three object elements, i.e. three column data of **Φ** (right).

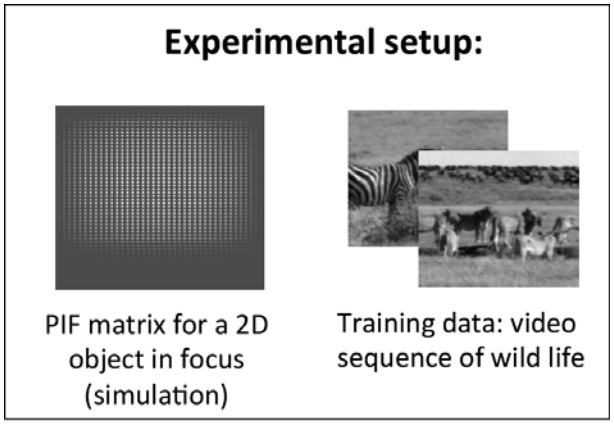


Fig.9 Results from training the dictionary for a plenoptic camera system.

In Fig.10 we show the reconstruction for the planar Doll object using the proposed dictionary learning algorithm, when placed in front of the plenoptic system. We have divided the original object into blocks, according to the field of view of each lenslet in an 11x11 array, and simulated the superpixel at the sensor behind each lenslet. White Gaussian noise of SNR = 60dB has been added to the sensor data. Object reconstructed using non-linear least square fitting, as proposed in 9). Using our proposed algorithm (Fig.4), we have estimated the sparse coefficient vectors **C** and the reconstructed blocks as **X** = Φ**C** shown in Fig.10. Peak Signal to Noise Ratio (PSNR) for the reconstructions using non-linear least square fitting is 22.5dB, while for the reconstruction using the dictionary learning method is 26.1dB. The visual quality is also better as shown in close-up images in Fig.10. One can notice some blocking artifacts due to per-lenslet processing of the sensor data. These can be removed by performing object reconstruction from 3x3 lenslet areas and then averaging. More details on reconstruction performance are discussed in 12).

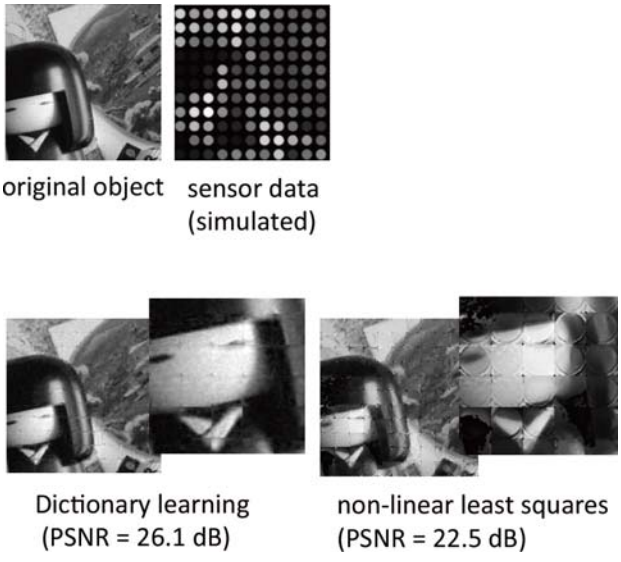


Fig.10 Spatial image reconstruction results using trained dictionary from simulated sensor data.

## 4. Conclusions

We developed an end-to-end system model for a plenoptic camera that enables evaluation of different imaging modalities, such as spectral and spatial scene information, allowing use of detailed optics and sensor models, and image processing algorithms. The model is used to design novel reconstruction methods for spectral and spatial information recovery. Experimental results for information recovery using software simulation and prototype data are presented.

### References


1) R. Ng et al.: Light field photography with a hand-held plenoptic camera, *Tech. Report*, Stanford University (2005).
2) T. Georgiev, G. Chunev, A. Lumsdaine: Superresolution with the Focused Plenoptic Camera, *Proc. IS&T/SPIE Electronic Imaging*, (2011).
3) R. Horstmeyer et al.: Flexible multimodal camera using a light field architecture, *IEEE Conf. Computational Photography* (2009).
4) A. Levin et al.: Understanding camera trade-offs through a Bayesian analysis of light field projections, *Proc. of the European Conference on Computer Vision (ECCV)* (2008).
5) T. E. Bishop, S. Zanetti, P. Favaro: Light Field Superresolution, *IEEE Conf. Computational Photography* (2009).
6) L. Meng, K. Berkner: System model and performance evaluation of spectrally coded plenoptic camera, *OSA Imaging Systems and Applications*, JW1A.3 (2012).
7) D. B. Cavanaugh et al.: VNIR hypersensor camera system, *Proc. SPIE Imaging Spectrometry XIV* 7457(1), 74570O (2009).
8) L. Meng, et al.: Evaluation of multispectral plenoptic camera, *Proceedings of SPIE*, Vol.8660 (2013).
9) S. A. Shroff, K. Berkner: High resolution image reconstruction for plenoptic imaging systems using system response, *OSA Computational Optical Sensing and Imaging*, CM2B.2 (2012).
10) R. G. Baraniuk et al., editors: Special issue on applications of sparse representation and compressive sensing, *Proc. IEEE*, 98(6) (2010).
11) E. J. Candes, J. Romberg: T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Communications on Pure and Applied Mathematics*, vol.59, no.8, pp.1207–1223 (2006).
12) D. L. Donoho, M. Elad: Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization, *Proceedings of the National Academy of Sciences*, vol.100, no.5, pp.2197–2202 (2003).
13) I. Tosic, S. A. Shroff, K. Berkner: Dictionary learning for incoherent sampling with application to plenoptic imaging, *Proc. of ICASSP* (2013).
14) I. Tosic, P. Frossard: Dictionary learning, *Signal Processing Magazine, IEEE*, vol.28, no.2, pp.27–38 (2011).