

LIGHT FIELD ACQUISITION FROM BLURRED OBSERVATIONS USING A PROGRAMMABLE CODED APERTURE CAMERA

P. Ruiz^a, J. Mateos^a, M. C. Cárdenas^b, S. Nakajima^c, R. Molina^a and A. K. Katsaggelos^d

(a) Dept. de Ciencias de la Computación e I.A., Universidad de Granada, Granada, Spain

(b) Instituto de Astrofísica de Andalucía, Granada, Spain*

(c) Optical Research Laboratory, Nikon Corporation, Japan

(d) Dept. of Electrical Engineering & Computer Science, Northwestern University, USA

ABSTRACT

In this paper we deal with the problem of acquiring a scene light field using a programmable coded aperture camera when the angular observations are out-of-focus. We describe a portable programmable coded aperture prototype that can be attached to any DSLR camera lens and propose a blind deconvolution method to deblur light fields. The performance of the proposed method is evaluated on synthetic and real images.

Index Terms— Computational photography, light field, blurred observations, programmable coded aperture camera.

1. INTRODUCTION

Moving from analog to digital has been a major advance in the world of photography. Besides the cost reduction, digital images can be edited and post-processed in countless ways by using a computer. In computational photography (CP), the postprocessing does most of the work, considering the image captured by the sensor as an intermediate data [1].

In the present work, we will use CP techniques to capture the light field of a scene. In recent years a number of light-field cameras have been developed. Plenoptic cameras, like Lytro [2] or Raytrix [3], introduce an array of microlenses in front of the sensor. This allows the sensor to record different angular views of the scene. Depending on the number of microlenses used, the resolution of the captured images can be greatly reduced. That is, there is a trade-off between angular resolution and spatial resolution of the light field; the more angular views are generated, the smaller the spatial resolution of each view.

To deal with this problem, systems using a coded aperture have been designed. In coded aperture acquisition systems, a

pattern mask is introduced to modify the lens aperture and to capture images that, once processed, allow to reconstruct the light field. Coded aperture began to be used for light field acquisition only a few years ago. In [4], the N angular views are obtained from N scrambled images captured with different masks and then solving a determined system of linear equations. The masks are loaded into a programmable LCD that is placed into the lens. Babacan et al. [5] reduce the number of observations required using Compressive Sensing theory in a system that uses an LCD to place the masks in front of the lens. The design by Nagahara [6] uses Liquid Crystal on Silicon (LCoS) to create the masks. This reduces the loss of light and improves the brightness and contrast but makes the lens bulkier than the LCD design.

None of the proposed models has dealt with the problem of defocused light fields. In spite of the small size of the individual blocks composing the coded aperture, the depth of field is limited and objects outside it will appear defocused in the reconstructed views. In this paper, we deal with the problem of blurred light field captured by the new coded aperture LCD based prototype, described in section 2, based on the design in [5], that can be mounted as a filter on any DSLR camera. To recover the light field from a set of blurred multiplexed observations, in section 3, we propose a new blind light field deconvolution method that adapts the model in [4] and the blind deconvolution method in [7] to our problem. The proposed method is evaluated on synthetic and real images and its performance is analyzed in section 4. Finally, section 5 concludes the paper.

2. PROTOTYPE DESCRIPTION

The coded aperture LCD based prototype we have constructed, see Fig. 1, can be mounted in front of the lens and has a small battery and controls so that it is portable and can be used autonomously. It uses an LCD array (Electronic Assembly DOGXL160S-7) consisting of 160×104 pixels of 0.418×0.397 mm with an active area of 70.0 mm \times 43.5 mm. In the prototype we have used a central part of 42

This research was supported by the Spanish Ministry of Economy and Competitiveness under project TIN2010-15137, the European Regional Development Fund (FEDER), and in part by the US Department of Energy grant DE-NA0000457.

Work in collaboration with the CP team members at IAA: J. Rodríguez Gómez, I. Bustamante Díaz, G. P. Candini, L. Costillo Iciarra, J. M. Jerónimo Zafra and M. R. Sanz Mesa.



Fig. 1. (a) Mechanical interface, electronic control board and LCD, mounted in the prototype. It is equipped on a Nikon D5000 camera with a Nikkor 50mm f/1.8 lens. (b) The LCD showing one of the coded apertures.

mm in diameter and baffled the remaining area of the LCD to minimize the stray light. A high level software has been developed in the Labview environment that automatically detects the connection of the prototype to the computer USB port. It also allows to create masks, load them from disk, store them locally or in the prototype, set the LCD contrast, and display a given mask stored in the prototype EEPROM. Also, a low level interface has been programmed in Matlab so that the prototype and the camera can be directly controlled from a PC. This simplifies the capture of pictures in batch mode.

The LCD allows four different transmission levels for each pixels: transparent, opaque and two intermediate gray levels. The transmission of the LCD has been measured in the visible spectral range, from 400 nm to 800 nm, in the four states (see Fig. 2) and the contrast of the LCD has been set to 95% in order to maximize the transmission when the pixel is “transparent” and to provide a good separation between the 2 gray states. Unfortunately, the transmission in the “opaque” state is not negligible, and the images have to be properly corrected. Furthermore, the images captured by the prototype suffer from a set of aberrations. Firstly, the LCD spectral transmittance is not uniform and, also, it is not the same at all spatial locations. Secondly, the location of the prototype with respect to the lens creates a mechanical vignetting effect that heavily affects apertures with diameter smaller than half the LCD size.

To ameliorate these problems, we concentrate on a small 30×30 pixels central part of the LCD where the transmittance of the LCD can be considered as spatially invariant. Also, we take into account only the central part of the images where no vignetting is present. This allows us to simplify the pre-processing of the captured images that, in fact, reduces to camera calibration. We only need to perform white balance using a white surface and take two calibration pictures of this surface; one with the LCD set to opaque and another to transparent. These images will allow us to recover the original luminance

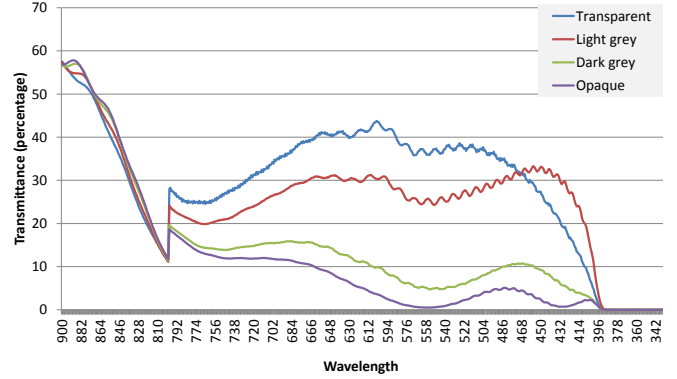


Fig. 2. LCD transmittance for the different wavelengths with a contrast of 95%.

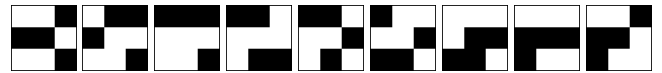


Fig. 3. Set of coded apertures used in the experiments. White means transparent and black corresponds to opaque.

of the scene despite the lower transmittance of the LCD.

3. IMAGE MODEL AND RECONSTRUCTION

By setting different blocks of the LCD to opaque or transparent we can capture different angular views of the same scene [4, 5, 6]. Opening only one block at a time allows to capture the light field sequentially by acquiring N angular views in N exposures. However, better results are obtained [4] using a multiplexed strategy where several blocks are set to transparent at the same time using a so called coded aperture.

To recover N views of the light field, we consider capturing M different pictures with coded apertures like the ones shown in Fig. 3. Then, each acquired image, \mathbf{y}_i , $i = 1, \dots, M$, is modeled as a linear combination of the different N , possibly blurred, angular views, as

$$\mathbf{y}_i = \sum_{j=1}^N a_{ij} \mathbf{H} \mathbf{x}_j + \mathbf{r}_i, \quad i = 1, \dots, M, \quad (1)$$

where \mathbf{x}_j is the j -th original (unknown) angular view of size $P = P_x \times P_y$ pixels, represented as a column vector. We assume that all the angular views share the same blur, \mathbf{H} , that is a $P \times P$ blurring matrix obtained from the unknown blur kernel \mathbf{h} of support $K = K_x \times K_y$, and \mathbf{r}_i is the capture noise. The a_{ij} coefficients indicate the contribution of the light field angular view j to picture i . Notice that, if the LCD behaved ideally, those coefficients would be 0 if the corresponding block in the LCD is set to opaque, or 1 if it is set to transparent. Unfortunately this is not the case on real LCDs but the values for the a_{ij} coefficients, with values between 0 and 1, can be estimated from the calibration pictures.

Our goal is to estimate the light field angular views $\mathbf{x}_i, i = 1, \dots, N$, and the blurring kernel, \mathbf{h} , from the set of $M = N$ multiplexed observed images $\mathbf{y}_i, i = 1, \dots, N$.

We first recover each pixel k of the different blurred angular views, represented by $\mathbf{z}_j, j = 1, \dots, N$, from the acquired images. Since the observations $\mathbf{y}_j, j = 1, \dots, N$, are noisy we utilize \mathbf{y}'_j , the denoised version of \mathbf{y}_j obtained by applying the BM3D [8] denoising method to the observed images before recovering the different blurred angular views. Then we obtain \mathbf{z}_j by solving the determined linear systems

$$\mathbf{y}'(k) = \mathbf{A}\mathbf{z}(k), \quad k = 1, \dots, P, \quad (2)$$

where the matrix \mathbf{A} is the $N \times N$ system matrix formed from the coefficients a_{ij} , where each row of the matrix contains the coefficients of a coded aperture and $\mathbf{z}(k)$ and $\mathbf{y}'(k)$ are column vectors formed by stacking the pixels at position k of the set of images $\{\mathbf{z}_1, \dots, \mathbf{z}_N\}$ and $\{\mathbf{y}'_1, \dots, \mathbf{y}'_N\}$, respectively.

According to our model each blurred angular view, $\mathbf{z}_j, j = 1, \dots, N$, can be mathematically expressed by

$$\mathbf{z}_j = \mathbf{H}\mathbf{x}_j + \mathbf{n}_j, \quad (3)$$

where the vector \mathbf{n}_j represents the noise, assumed to be Gaussian of variance β^{-1} . Its precision parameter, β , is the same for all the images because, as they are taken under identical conditions, they will have the same noise properties. Notice that \mathbf{n}_j was introduced since \mathbf{z}_j of Eq. (2) will very likely be noisy.

We apply the variational Bayesian approach in a blind deconvolution procedure [7] to recover the blurring kernel \mathbf{h} and the restored angular views \mathbf{x}_j . From Eq. (3), we write the degradation model as

$$\begin{aligned} p(\mathbf{z}|\mathbf{x}, \mathbf{h}, \beta) &= \prod_{j=1}^N p(\mathbf{z}_j|\mathbf{x}_j, \mathbf{h}, \beta) \\ &\propto \beta^{P/2} \exp\left(-\frac{\beta}{2} \sum_{j=1}^N \|\mathbf{z}_j - \mathbf{H}\mathbf{x}_j\|^2\right), \end{aligned} \quad (4)$$

where \mathbf{z} and \mathbf{x} are column vectors formed by stacking vertically the vectors \mathbf{z}_j and $\mathbf{x}_j, j = 1, \dots, N$, respectively.

We use the general TV function as image prior for each view and, hence, we define

$$p(\mathbf{x}) = \prod_{j=1}^N p(\mathbf{x}_j|\alpha) \propto \exp\left(-\alpha \sum_{j=1}^N \text{TV}(\mathbf{x}_j)\right), \quad (5)$$

where

$$\text{TV}(\mathbf{x}_j) = \sum_{k=1}^P \sqrt{(\Delta^h(\mathbf{x}_j)(k))^2 + (\Delta^v(\mathbf{x}_j)(k))^2}, \quad (6)$$

with the operators $\Delta^h(\mathbf{x}_j)(k)$ and $\Delta^v(\mathbf{x}_j)(k)$ corresponding to the horizontal and vertical first order differences at pixel k , respectively.

To estimate all unknowns $\Theta = \{\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{h}\}$ the variational Bayesian approach is used. In this approach, the posterior $p(\Theta|\mathbf{z})$ is approximated by another distribution, $q(\Theta)$, by minimizing the Kulback-Leibler (KL) divergence between both distributions [9]. A convenient factorization of $q(\Theta) = q(\mathbf{x}_1) \dots q(\mathbf{x}_N)q(\mathbf{h})$, named mean field approximation [9], is used in order to get a tractable minimization problem.

3.1. Angular view estimation

Due to use of TV prior, for estimating the distribution of each angular view, $q(\mathbf{x}_j)$, it is necessary to carry out a majorization-minimization procedure, as described in [7]. Thus, $q(\mathbf{x}_j)$ is estimated as a Gaussian distribution with mean $\bar{\mathbf{x}}_j$ and covariance matrix $\Sigma_{\mathbf{x}_j}$ given by

$$\bar{\mathbf{x}}_j = \Sigma_{\mathbf{x}_j} \beta \bar{\mathbf{H}}^T \mathbf{z}_j \quad (7)$$

$$\Sigma_{\mathbf{x}_j} = (\beta \bar{\mathbf{H}}^T \bar{\mathbf{H}} + \alpha((\Delta^h)^T \mathbf{W}_j \Delta^h + (\Delta^v)^T \mathbf{W}_j \Delta^v))^{-1} \quad (8)$$

where $\bar{\mathbf{H}}$ is the convolution matrix obtained from the current estimation of \mathbf{h} , $\bar{\mathbf{h}}$, and $\mathbf{W}_j = \text{diag}((u_j(k))^{-1/2}), k = 1, \dots, P$, with $u_j(k)$ a set of additional parameters introduced in the majorization procedure and calculated [7] as

$$u_j(k) = (\Delta^h(\mathbf{x}_j)(k))^2 + (\Delta^v(\mathbf{x}_j)(k))^2. \quad (9)$$

Notice that $\Sigma_{\mathbf{x}_j}$ in Eq. (8) is a $P \times P$ matrix and therefore its computation is extremely expensive. To alleviate this problem, each restored view, $\bar{\mathbf{x}}_j$, is estimated by solving, using conjugate gradient, the linear equation system

$$(\beta \bar{\mathbf{H}}^T \bar{\mathbf{H}} + \alpha((\Delta^h)^T \mathbf{W}_j \Delta^h + (\Delta^v)^T \mathbf{W}_j \Delta^v)) \mathbf{x}_j = \beta \bar{\mathbf{H}}^T \mathbf{z}_j. \quad (10)$$

3.2. Blur estimation

Note that Eq. (3) can also be written as $\mathbf{z}_j = \mathbf{X}_j \mathbf{h} + \mathbf{n}_j$ by forming the matrix \mathbf{X}_j similarly to \mathbf{H} . To estimate the blur, we follow the approximation proposed in [10] where \mathbf{h} is assumed to have a degenerate distribution $q(\mathbf{h})$ and the value where the distribution is degenerate is calculated as the PSF solution of

$$\hat{\mathbf{h}} = \arg \min_{\mathbf{h}} \mathbb{E}[\beta \sum_{i=1}^N \|\mathbf{z}_i - \mathbf{H}\mathbf{x}_i\|^2]. \quad (11)$$

Let us approximate \mathbf{W}_j in Eq. (8), following [7], by $\mathbf{W}_j \approx \text{mean}(\text{diag}(\mathbf{W}_j) \mathbf{I}_{P \times P})$, and then, following [10], utilize

$$\Sigma_{\mathbf{x}_j} \approx s_{\mathbf{x}_j} \mathbf{I}_{P \times P}, \quad (12)$$

with $s_{\mathbf{x}_j} = (\beta \sum_{k=1}^K \mathbf{h}(k)^2 + 4\alpha \text{mean}(\text{diag}(\mathbf{W}_j)))^{-1}$. Let

$$\mathbf{C}_{\mathbf{h}}^{-1} = \sum_{j=1}^N (\bar{\mathbf{X}}_j^T \bar{\mathbf{X}}_j + P s_{\mathbf{x}_j} \mathbf{I}_{K \times K}), \quad (13)$$

with $\bar{\mathbf{X}}_j$ the convolution matrix obtained from the current estimation of $\mathbf{x}_j, \bar{\mathbf{x}}_j$. Then, $\hat{\mathbf{h}}$ can be approximated as the solution of the restricted quadratic program problem

$$\begin{aligned} \hat{\mathbf{h}} &= \arg \min_{\mathbf{h}} \mathbf{h}^T \mathbf{b}_h + \frac{1}{2} \mathbf{h}^T \mathbf{C}_h^{-1} \mathbf{h}, \\ \text{subject to } \sum_{k=1}^K \mathbf{h}(k) &= 1, \\ \mathbf{h}(k) &\geq 0, \quad k = 1, \dots, K. \end{aligned} \quad (14)$$

with

$$\mathbf{b}_h = - \sum_{j=1}^N \bar{\mathbf{X}}_j^T \bar{\mathbf{z}}_j. \quad (15)$$

In summary, to recover the light field from the degraded observations we proceeded as follows. First, we denoise the observed images by applying the BM3D [8] denoising method and then recover the different blurred angular views from Eq. (2). Secondly, we estimate the blur from the luminance band of the blurred views by alternatively iterating between Eqs. (10) and (14). The rationale behind this process is that the blur contaminating the R, G, and B bands is the same since these bands were captured under the same conditions and so we can speed up the estimation process by using only the luminance band. Finally, once the blur is obtained, we estimate each one of the RGB bands of the restored angular views by applying the non-blind restoration procedure described by Eq. (10) with the already estimated blur.

4. EXPERIMENTAL RESULTS

We have evaluated the performance of the proposed method with synthetic and real images. In the synthetic experiment, a scene was created with Blender¹ and a set of 9 different angular views were taken by placing a pinhole camera at 9 coplanar positions in the space. Those positions formed a 3×3 grid in a plane perpendicular to the Z scene axis. The angular view at position 5 (center) of the grid is displayed in Fig. 4a.

We then generated the set of 9 coded apertures depicted in Fig. 3 by selecting random 3×3 binary masks that have 5 open blocks. Each single block was in total open the same number of times in the 9 coded apertures set. The observed set of images was obtained by simulating the capture process in Eq. (1), that is, first blurring each view with a Gaussian blur with variance 1 and then multiplexing the blurred views using the set of coded apertures shown in Fig. 3. Finally, Gaussian noise with standard deviation 0.001 was added to obtain the observed images, whose observation number 5 is depicted in Fig. 4b. Note that the letters in ‘‘Hola’’ that are at the focal plane are only blurred (since they will be at the same position

¹Available at <http://www.blender.org/>

Table 1. Mean PSNR and SSIM for the R,G,B bands and the mean of the RGB images for the synthetic experiment.

	R	G	B	mean (RGB)
PSNR	37.10	35.15	34.35	35.53
SSIM	0.9888	0.9868	0.9873	0.9876

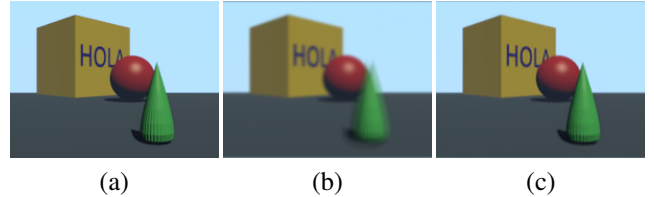


Fig. 4. Synthetic experiment with Gaussian blur ($\sigma = 1$): (a) Original angular view 5, (b) Simulated captured image with mask 5 in Fig. 3, (c) reconstructed angular view 5.

in all the views) while the cone, that is far from the focal plane, represents the mixture of the different blurred views.

To estimate the original angular views from the observed images we apply the reconstruction algorithm described in the previous section. The initial blur \mathbf{h}^0 is set to a Gaussian with variance 0.16 and support $K_x = K_y = 21$, hence $K = 441$, that is a PSF close to a delta function. The precision parameter β in Eq. (3) is chosen such that the value of $P_{S_{x_k}}$ in Eq. (13) is a fraction (0.1) of the maximum value of $\mathbf{X}^t \mathbf{X}$ in the first iteration of the algorithm. The rationale behind this is that the value of $P_{S_{x_k}} \mathbf{I}_{K \times K}$, that represents the uncertainty of the minimum squares solution, tends to be smaller as we are more certain on the value of the image so, in the first iterations, we are forcing some uncertainty in the blur estimation process that will be reduced as the image is better restored. The image prior parameter α is chosen as a fraction of the value of β . We chose $\alpha = 0.001\beta$ to preserve most of the original data while smoothing out the restoration artifacts and the noise. The estimated angular view 5 is presented in Fig. 4c. Note that the blur has been successfully removed while preserving the structure in the cone. Numerical results, shown in Table 1, show that the reconstructed images have a very high quality both in terms of PSNR and SSIM measures.

We also tested the proposed method on real images. The set of images was taken with the prototype using the set of coded apertures depicted in Fig. 3. To minimize the effects of the spatially variant degradations produced by the LCD, we concentrated on a square of 30×30 pixels in the center of the LCD which was divided in a 3×3 set of square apertures each of size 10×10 pixels. This means that the area of each single block is 16.6 mm^2 . Also, we used only the 512×512 pixel central part of the images to reduce the spatially variant effects of the lens and prevent vignetting from appearing.

The scene, as seen in Fig. 5, was set at 800 mm from the camera, the distance from the pin to the background is 40 mm, and, when 5 blocks are open, the depth of field is 45.1 mm.

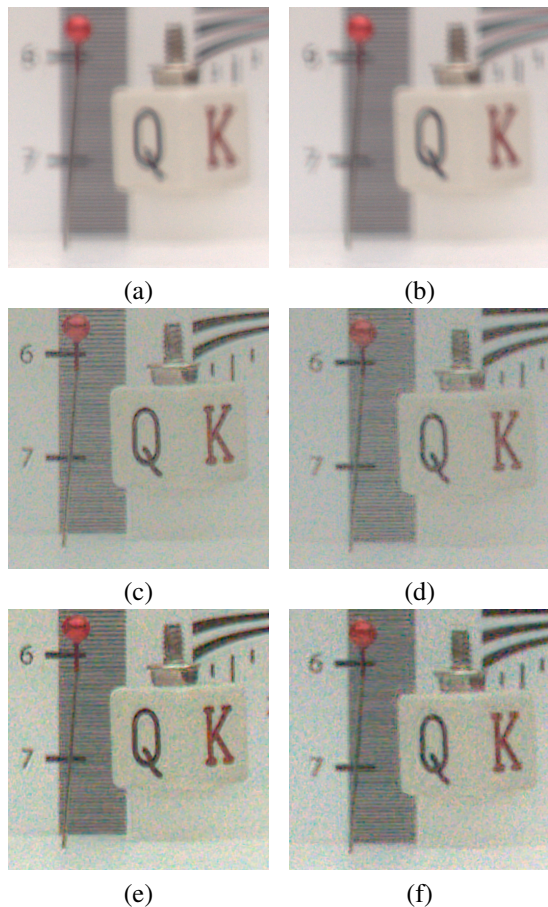


Fig. 5. Real experiment focused at 50 mm from the center of the scene. (a) Observed image 1, (b) Observed image 9, (c) Demultiplexed blurred angular view 1, (d) Demultiplexed blurred angular view 9, (e) deblurred angular view 1, (f) deblurred angular view 9.

We took pictures focusing at 50 mm from the dice (see Figs. 5a and 5b). For each RGB band, the system matrix \mathbf{A} was obtained by setting its coefficients equal to the mean value of the calibration pictures with the LCD set to opaque or to transparent, depending on whether the corresponding block is opaque or transparent. This allows us to recover the blurred angular views without any additional preprocessing.

We applied BM3D to the observed images using a variance calculated from a flat region of the image. Then we recovered the different blurred angular views from Eq. (2) resulting in the images depicted in Figs. 5c and 5d. Finally, the deconvolution algorithm was applied to the blurred angular views following the procedure described for synthetic images, obtaining the restored views, two of which are shown in Figs. 5e and 5f. As it can be observed, the restored views are sharp, making clearly visible the lines in the background or the details in the thread on the screw, but a bit noisy. This is due to noise amplification in the demultiplexing stage.

5. CONCLUSIONS

We have presented a new programmable aperture camera prototype that allows to capture light fields. We have addressed the problem of recovering blurred light fields that may occur due to the limited depth of field of the cameras. We have developed a method for deconvolving those blurred light fields and tested it on both synthetic and real images.

6. REFERENCES

- [1] M. Levoy, “Experimental platforms for computational photography,” *IEEE Computer Graphics and Applications*, vol. 30, pp. 81–87, 2010.
- [2] Y-R Ng, C. Pitts, and T. Knight, “Light field data acquisition,” U.S. Patent Application 20120327222, 2012.
- [3] C. Perwass and L. Wietzke, “Light field camera technology,” <http://www.raytrix.de/index.php/Technology.html>, Mar. 2013.
- [4] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen, “Programmable aperture photography: multiplexed light field acquisition,” in *ACM SIGGRAPH*, 2008, pp. 55:1–55:10.
- [5] S. D. Babacan, R. Ansorge, M. Luessi, P. Ruiz, R. Molina, and A. K. Katsaggelos, “Compressive light field sensing,” *IEEE Trans. on Image Processing*, vol. 60, pp. 3964–3977, 2012.
- [6] H. Nagahara, C. Zhou, T. Watanabe, H. Ishiguro, and S. K. Nayar, “Programmable aperture camera using LCoS,” in *Proc. of the ECCV’10*, 2010, pp. 337–350.
- [7] S. D. Babacan, R. Molina, and A. K. Katsaggelos, “Variational Bayesian blind deconvolution using a total variation prior,” *IEEE Trans. on Image Processing*, vol. 18, pp. 12–26, 2009.
- [8] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3D transform-domain collaborative filtering,” *IEEE Trans. on Image Processing*, vol. 16, pp. 2080–2095, 2007.
- [9] C. M. Bishop, *Pattern Recognition and Machine Learning*, Information Science and Statistics. Springer, 2007.
- [10] S. D. Babacan, R. Molina, M. Do, and A. K. Katsaggelos, “Blind deconvolution with general sparse image priors,” in *Proc. of the ECCV’12*, 2012, pp. 341–355.