Compressive Light Field Sensing

S. Derin Babacan, *Member, IEEE*, Reto Ansorge, Martin Luessi, *Member, IEEE*,
Pablo Ruiz Matarán, *Student Member, IEEE*, Rafael Molina, *Member, IEEE*,
and Aggelos K. Katsaggelos, *Fellow, IEEE*

Abstract—We propose a novel design for light field image acquisition based on compressive sensing principles. By placing a randomly coded mask at the aperture of a camera, incoherent measurements of the light passing through different parts of the lens are encoded in the captured images. Each captured image is a random linear combination of different angular views of a scene. The encoded images are then used to recover the original light field image via a novel Bayesian reconstruction algorithm. Using the principles of compressive sensing, we show that light field images with a large number of angular views can be recovered from only a few acquisitions. Moreover, the proposed acquisition and recovery method provides light field images with high spatial resolution and signal-to-noise-ratio, and therefore is not affected by limitations common to existing light field camera designs. We present a prototype camera design based on the proposed framework by modifying a regular digital camera. Finally, we demonstrate the effectiveness of the proposed system using experimental results with both synthetic and real images.

Index Terms—Bayesian methods, coded aperture, compressive sensing, computational photography, image reconstruction, light fields.

I. INTRODUCTION

RECENT advances in computational photography [1] provided effective solutions to a number of photographic problems, and also resulted in novel methods to acquire and process images. Novel camera designs allow for the capturing of information of the scene which is not possible to obtain using traditional cameras. This information can then be used for example to generate the three-dimensional scene geometry, or for novel applications, such as digital refocusing or synthetic aperture [1].

Light field cameras are one of the most widely used class of computational cameras. The light field expresses the radiance density function on the camera sensor, or the light energy of all rays in 3D space passing through the camera. For instance, a four-dimensional (4D) discrete light field image $\mathbf{x}(i, k, m, n)$ with spatial dimensions i, k and angular dimensions m, n contains images of a scene from a number of angles, which provide information about the 3D structure of the scene. Each 2D image $\mathbf{x}(i, k, m_0, n_0)$ with fixed angular coordinates m_0, n_0 is called an *angular image*. Traditional cameras integrate these angular images (or equivalently, light rays) over their 2D aperture to obtain the image, which results in the loss of valuable depth information about the scene. On the other hand, light field cameras capture the angular data and provide means to work directly with the light-rays instead of pixels, allowing one to produce many views of the scene, or perform many photographic tasks after the acquisition is made. This provides a clear advantage for light field imaging over traditional photography and makes many novel applications possible.

Compressive sensing (CS) [3], [4] has recently become very popular due to its interesting theoretical nature and wide area of applications. The theory of compressive sensing dictates that a signal can be recovered very accurately from a much smaller number of measurements than required by traditional methods, provided that it is *compressible* (or *sparse*) in some transform basis, i.e., only a few basis coefficients contain the major part of the signal energy. Besides sparsity, compressive sensing makes use of the incoherent measurement principle[1], and has led to many interesting theoretical results and novel applications (see, for instance, [5], [6]).

In this paper, we present a novel application of compressive sensing, namely, a novel framework to acquire light field images. We show that light field acquisition can be formulated using a incoherent measurement principle. We then demonstrate that light field images have a highly sparse nature, which, in combination with incoherent measurements, can be exploited to reconstruct the light field images with much fewer image acquisitions than traditionally required. By exploiting this sparsity in light field images, we develop a novel reconstruction algorithm that recovers the original images from few compressive measurements with a very high degree of fidelity.

S. D. Babacan is with the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: dbabacan@illinois.edu).

R. Ansorge is with Varian Medical Systems, Baden CH-5405, Switzerland (e-mail: reto.ansorge@gmx.ch).

M. Luessi is with the Department of Radiology, Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114 USA (e-mail: mluessi@nmr.mgh.harvard.edu).

P. R. Matarán and R. Molina are with the Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Granada 18010, Spain (e-mail: mataran@decsai.ugr.es; rms@decsai.ugr.es).

A. K. Katsaggelos is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208-3113 USA (e-mail: aggk@eecs.northwestern.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2012.2210237

[1]Loosely speaking, incoherent measurements refer to non-adaptive and uncorrelated with the signal of interest. See, for instance, [3], [5] for technical definition and interpretations.

In addition, we propose a novel camera design based on the developed acquisition framework. We build our design on ideas from coded aperture imaging, computational photography and compressive sensing. By exploiting the fact that different regions of the aperture of a camera correspond to images of the scene from different angles, we incorporate a compressively coded mask placed at the aperture to obtain incoherent measurements of the incident light field. These measurements are then decoded using the proposed reconstruction algorithm to recover the original light field image. We exploit the highly sparse nature of the light field images to obtain accurate reconstructions with a small number of measurements compared to the high angular dimension of the light field image. The proposed camera design provides images with high signal-to-noise ratio and does not suffer from the spatio-angular resolution trade-off in most existing light field camera designs. Finally, we demonstrate the efficiency of the proposed framework with both synthetic experiments and real images captured by a prototype camera.

The paper is organized as follows. First we review related prior work in light field and coded aperture imaging in Sec. II. In Sec. III we present the proposed acquisition method to obtain incoherent measurements of the light field image. We model the acquisition system and the light field images using a Bayesian framework, which is described in Sec. IV. The Bayesian inference procedure used to develop the reconstruction algorithm is presented in Sec. V. We present a prototype light field camera based on the proposed framework in Sec. VI. The effectiveness of the proposed system is demonstrated with both synthetic and real light field images in Sec. VII and conclusions are drawn in Sec. VIII.

II. Related Prior Work

A. Light Field Acquisition

Light field acquisition, based on the principles of integral photography, was first proposed over a century ago [7], [8]. The same ideas appeared in the computer vision literature first as the *plenoptic camera* [9], and then the potential of light field imaging was demonstrated in [10] and [11]. The original design in [9] is implemented in a hand-held camera in [12], where a microlens (lenticular) array is placed between the main lens and the camera sensor. A similar approach is proposed in [13], where instead of using microlenses, a lens array is placed in front of the camera main lens. In both approaches, the light field image is captured using one acquisition. The additional lens array is used to capture the angular information, and reordering the captured image results in images of different views of the scene. Other proposed light field camera designs include multi-camera systems [14] and mask-based designs [15], [16], which encode the angular information using frequency-multiplexing.

Many of these designs suffer from the spatio-angular resolution[2] trade-off [13], that is, one cannot obtain light field images with both high spatial- and high angular resolution.

[2]The spatial and angular resolution here only refer to the number of digitally acquired elements such as pixels and images. Certain optical effects such as diffraction due to aperture size are not included in this analysis.

This problem is inherent in designs with one recording sensor (or film) and where only one acquisition is made. If the captured light field image has an angular resolution of $N_h \times N_v$, and a spatial resolution of $P_h \times P_v$, then $N_h \times N_v \times P_h \times P_v$ can only be less than or equal to the number of pixels in the camera sensor. For instance, a typical light field image captured using the plenoptic camera in [12] provides 14×14 angular images of size approximately 300×300 in a 16 megapixel camera. Multi-camera systems [14] are not affected from the spatio-angular resolution trade-off, but they are very costly to implement and cumbersome for practical usage.

Recently, a *programmable aperture camera* is proposed [17], where a binary mask is used to code the aperture. Angular images are *multiplexed* into single 2D images similarly to the principle of coded aperture imaging. After multiple acquisitions are made, a linear estimation procedure is employed to recover the full light field image. Although this design captures images with both high spatial and angular resolution, the number of acquisitions are equal to the number of angular dimensions. Therefore, obtaining a light field image with a high angular resolution is not practical.

During the development of this work, we became aware of [18], which appeared after [2], and independently considered the application of compressive sensing to light field acquisition. The work in [18] devises a linear recovery procedure from compressive measurements incorporating statistical correlations among the angular images via their autocorrelation matrix. In contrast, in this work we exploit the structure within the light field more explicitly using nonlinear relationships among the angular images.

B. Coded Aperture Imaging

Coded aperture imaging is developed in order to collect more light in situations where a lens system cannot be used, due to the measured wavelengths. Imaging systems with coded apertures are currently widely used in astronomy and medicine. The technique is based on the principle of pinhole cameras, but instead of only one pinhole which suffers from low SNR, a specially designed array of pinholes is used. This array of pinholes provides images that are overlapping copies of the original scene, which can then be decoded using computational algorithms to provide a sharp image. There is a vast literature on coded aperture methods in astronomy and medicine (see, for example, [19], [20]).

Recent works considered coded aperture methods for developing novel image acquisition methods. In [21], the aperture is coded in the time-domain to modify the exposure for motion deblurring. Spatially modifying the aperture has been used for a range of applications: Levin *et. al.* [22] proposed utilizing an aperture mask to reconstruct both the original image and the depth of the scene from a single snapshot. A lensless imaging system is proposed in [23] that allows for the manipulation of the captured scene in ways not possible by traditional cameras, such as splitting field of view. Nayar *et. al.* [24] used a spatial light modulator to control the exposure per pixel, which can be used to obtain high-dynamic range images. Other uses
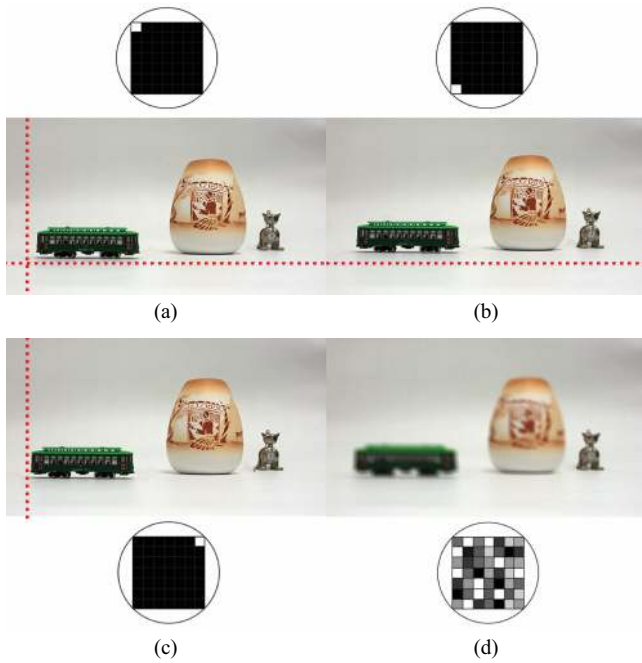
Fig. 1. Basic principle of utilizing a coded aperture to obtain light field images. (a)–(c) The angular images when only corner blocks of the aperture are left open. Both horizontal and vertical parallax can be observed between these images (horizontal and vertical dashed lines are shown to denote the vertical and horizontal parallax, respectively). (d) Captured image with the randomly coded aperture used in the proposed compressive sensing light field camera. All images are from a synthetic light field image (see Section VII).

of coded apertures include super resolution [25] and range estimation [26].

Compressive sensing methods have also been applied in conjunction with coded apertures or compressively coded blocking masks. Novel imaging methods have been proposed for spectral imaging [27], dual-photography [28], and the design of structured light for recovering inhomogeneous participating media [29]. Most recently, compressively coded aperture masks are used for single-image super-resolution and shown to provide higher quality images than traditional coded apertures [30], [31].

A related approach to coded aperture imaging is *wavefront coding* [32], where the image is intentionally defocused using phase plates so that the defocus is uniform throughout the image. The captured image can then be deconvolved to obtain an image with an enlarged depth of field.

III. COMPRESSIVE SENSING OF LIGHT-FIELDS

In this section, we will show that light field image acquisition can be formulated within the compressive sensing framework. We first show that light field images can be acquired by coding the camera aperture, and then present the proposed compressive acquisition system. In the following, a 4D light field image is denoted by $\mathbf{x}$, which is the collection of N angular images $\mathbf{x}^j$, such that $\mathbf{x} = \{\mathbf{x}^j\}$, $j = 1, \ldots N$.

A. Light-Field Acquisition by Coded Apertures

A fundamental principle used in this work is that different regions of the aperture capture images of the scene from different angles[3] [17], [23], [33], [34]. Specifically, the main camera lens can be interpreted as an array of multiple virtual lenses (or cameras). This concept is illustrated in Fig. 1(a)-(c), where only certain parts (white blocks) of the aperture are left open. As can be seen from Fig. 1(a)-(c), the acquired images exhibit vertical and horizontal parallax. By separately opening one region of the aperture and blocking light in the others, the complete light field with an angular dimension of N can be captured with N exposures. However, obtaining the light field image in this fashion has two disadvantages: First, due to the very small amount of light arriving to the sensor at each exposure, the captured angular images have very low signal-to-noise ratios (SNR). Second, a large number of acquisitions have to be made in order to obtain high angular resolution. The programmable aperture camera design in [17] addressed the first problem by incorporating a multiplexing scheme, but the second problem remains a serious drawback.

We address both of these issues by using a randomly coded non-refractive mask in front of the aperture. Each image acquired in this fashion is a random linear combination (and therefore an incoherent measurement) of the angular images. An example image captured in this fashion is illustrated in Fig. 1(d), where the amount of light passing through different regions of the aperture are randomly selected (shown at the bottom of Fig. 1(d)). As shown in the following, using such a random mask overcomes both of the problems described above.

The mathematical principle behind this idea is formulated as follows. Let us assume that the aperture of the main camera lens is divided into N blocks, with $N = N_h \times N_v$ where N_h and N_v represent the number of horizontal and vertical divisions. During each acquisition i, each block j is assigned a weight $0 \le a^{ij} \le 1$ which controls the amount of light passing through this block. Therefore, a^{ij} represents the transmittance of the block j, i.e., it is the fraction of incident light that passes through the block. As mentioned above, each block j captures an angular image $\mathbf{x}^j$ in the light field image, and therefore the acquired image $\mathbf{y}^i$ at the ith acquisition can be represented as a linear combination of the N angular images as

$$\mathbf{y}^i = \sum_{j=1}^{N} a^{ij} \mathbf{x}^j, \quad i = 1, \ldots, M \quad (1)$$

where we use the vector notation such that $\mathbf{y}^i$ and $\mathbf{x}^j$ are both $P \times 1$ vectors, with P the number of pixels in each image. Note that in a traditional camera, the acquired image is the average of all angular images, i.e., $a^{ij} = \frac{1}{N}$, since the aperture integrates all light rays coming from different directions.

After M acquisitions ($M \le N$), the complete set of observed images $\{\mathbf{y}^i\}$ can be expressed in matrix-vector

[3]This is a widely used model employing a geometric optics perspective. A more recent analysis of lightfields based on wave optics provide additional views on the transformation of light fields through lenses [35].

form as

$$\begin{pmatrix} \mathbf{y}^1 \\ \mathbf{y}^2 \\ . \\ . \\ \mathbf{y}^M \end{pmatrix} = \begin{pmatrix} a^{11}\mathbf{I} & a^{12}\mathbf{I} & . & . & a^{1N}\mathbf{I} \\ a^{21}\mathbf{I} & a^{22}\mathbf{I} & . & . & a^{2N}\mathbf{I} \\ . & . & . & . & . \\ . & . & . & . & . \\ a^{M1}\mathbf{I} & a^{M2}\mathbf{I} & . & . & a^{MN}\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ . \\ . \\ \mathbf{x}^N \end{pmatrix} \qquad (2)$$

with $\mathbf{I}$ the $P \times P$ identity matrix. The system in (2) is expressed in a more compact form as

$$\mathbf{y} = \mathbf{A}\mathbf{x} \qquad (3)$$

with

$$\mathbf{A} = \begin{pmatrix} a^{11} & a^{12} & . & . & a^{1N} \\ a^{21} & a^{22} & . & . & a^{2N} \\ . & . & . & . & . \\ . & . & . & . & . \\ a^{M1} & a^{M2} & . & . & a^{MN} \end{pmatrix} \otimes \mathbf{I} = \hat{\mathbf{A}} \otimes \mathbf{I} \qquad (4)$$

where $\otimes$ is the Kronecker product. Taking also the acquisition noise into account, the final observation model can be expressed as

$$\mathbf{y} = \left(\hat{\mathbf{A}} \otimes \mathbf{I} \right) \mathbf{x} + \mathbf{n} = \mathbf{A}\mathbf{x} + \mathbf{n} \qquad (5)$$

with $\mathbf{n}$ the $PM \times 1$ noise vector.

B. Compressively Coded Apertures

If the linear measurement matrix $\mathbf{A}$ satisfies certain properties dictated by the theory of compressive sensing [3], the light field acquisition system in (5) can be seen as a noisy incoherent measurement system. A sufficient condition for a matrix to be a compressive sensing matrix is the *restricted isometry property* (RIP) [3], [36], which is proven to hold with a very high probability for a general class of matrices with their entries drawn from certain random probability distributions. For instance, if $\hat{\mathbf{A}}$ in (5) is constructed by independently drawing its entries from a Gaussian distribution, then $\hat{\mathbf{A}}$ satisfies RIP with an overwhelming probability.

It is straightforward to show that if $\hat{\mathbf{A}}$ is a valid compressive sensing matrix, then $\mathbf{A}$ is a valid compressive sensing matrix as well. A simple proof is as follows. The mutual coherence of matrix $\hat{\mathbf{A}}$ is given by [37]

$$\mu\left(\hat{\mathbf{A}}\right) = \max_{i \neq j} \frac{|\hat{A}_i^T \hat{A}_j|}{\| \hat{A}_i \| \| \hat{A}_j \|} \qquad (6)$$

where $\hat{A}_i$ is the ith column of $\hat{\mathbf{A}}$. The mutual coherence characterizes the correlation between the columns of matrix $\hat{\mathbf{A}}$, and it is always positive for matrices with more columns than rows. It is shown that the mutual coherence provides a bound for the RIP constants [38], and therefore RIP-based guarantees can be applied using mutual coherence. Using properties of the Kronecker product, it can be seen that

$$\mathbf{A}^T \mathbf{A} = \left(\hat{\mathbf{A}} \otimes \mathbf{I} \right)^T \left(\hat{\mathbf{A}} \otimes \mathbf{I} \right) \qquad (7)$$

$$= \hat{\mathbf{A}}^T \hat{\mathbf{A}} \otimes \mathbf{I}. \qquad (8)$$

Thus, the inner products of columns of $\mathbf{A}$ have the exact same values as the columns of $\hat{\mathbf{A}}$, and therefore they have

the same mutual incoherence. If the mutual coherence of $\hat{\mathbf{A}}$ is sufficiently small so as to satisfy RIP [38], $\mathbf{A}$ also satisfies the restricted isometry property and it is therefore a valid compressive sensing matrix.

Based on this, the acquisition system in (5) is an incoherent measurement system of angular images $\mathbf{x}^j$, where each acquired image is a random linear combination of the angular images. The theory of compressive sensing then dictates that if the unknown image $\mathbf{x}$ can be represented sparsely in some transform domain, then it can be recovered with much fewer measurements than traditionally required ($M \ll N$). Due to the nature of multi-view images and especially in the specific case considered in this work where the angular images are aligned on a small-baseline, the redundancy within the light field images is very high. In fact, there are multiple sources of sparsity inherent in light field images, due to correlations both within and in between the angular images (see Sec. IV-B for details). Therefore, light field images can be very accurately reconstructed with very few acquisitions by utilizing the compressive acquisition system in (5) and by exploiting their sparse nature within nonlinear reconstruction frameworks.

An important design issue is the selection of the measurement matrix $\mathbf{A}$, which determines the level of incoherence of the measurements and therefore the reconstruction performance. The design of measurement matrices for compressive sensing is an active area of research, and many of the existing designs can be used for the proposed aperture mask. In this work, we specifically experimented with two different types of measurement matrices, namely, uniform spherical and scrambled Hadamard ensembles [39]. If fractional values of the block transmittances are permitted, a general class of matrices can be utilized, with positivity of the matrix entries as the only constraint. In this case, uniform spherical ensembles (with values ranging between 0 and 1) are very suitable as measurement matrices. If the mask is limited to binary codes, scrambled Hadamard ensembles can be used to code the aperture. Moreover, the measurement matrices can also be selected depending on specific requirements of the optical systems, e.g., the expected amount of transmitted light can be varied by varying the mean value of the corresponding probability distribution, or by choosing a specific construction of the random measurement matrix.

It should be noted that since many (or possibly all) blocks are open in each exposure, each captured image has a high SNR due to the small amount of loss of light. In fact, the measurement matrices can be designed to optimize the amount of passing light while maintaining the random structure. Moreover, as shown in the experimental results section, incorporating a nonlinear reconstruction mechanism provides images with much higher SNRs than those of linear reconstruction methods, such as demultiplexing.

Finally, it should be noted that the coded aperture setup used in this work is a specific application of the acquisition system in (5). The proposed compressive sensing formulation for light field acquisition can be applied to a wider range of light field imaging applications. For instance, multiple camera or multiple lens imaging systems such as camera arrays

and stereo cameras can equally well incorporate the incoherent measurement system in (5) and significantly reduce the number of acquisitions without sacrificing spatial or angular resolution.

IV. HIERARCHICAL BAYESIAN MODEL FOR RECONSTRUCTION

In order to reconstruct the angular images $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^N$ from the incoherent measurements $\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^M$ and $\mathbf{A}$, both the observation process (5) and the unknown light field image $\mathbf{x}$ have to be modeled. For this modeling, we use the hierarchical Bayesian framework by employing a conditional distribution $p(\mathbf{y}|\mathbf{x}, \beta)$ for the observation model in (5) and a *prior* distribution $p(\mathbf{x}|\boldsymbol{\alpha}_{\mathrm{TV}}, \boldsymbol{\alpha}_c)$ on the unknown light field image $\mathbf{x}$. These distributions depend on the model parameters β, $\boldsymbol{\alpha}_{\mathrm{TV}}$ and $\boldsymbol{\alpha}_c$, which are called *hyperparameters*. In the second stage of the hierarchical model we use additional prior distributions, called *hyperpriors*, to model them. In the following subsections, we present the specific forms of each of these distributions.

A. Observation (Noise) Model

The observation noise is assumed to be independent and Gaussian with zero mean and variance equal to β^{-1}, that is, using (5),

$$p(\mathbf{y}|\mathbf{x}, \beta) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x}, \beta^{-1}). \qquad (9)$$

B. Light-Field Image Model

The choice of randomly programmed coded apertures makes the exact/approximate recovery of the angular images possible through the use of sparsity inherent in light field images. There are two sources of sparsity within a light field image that can be exploited. The first one is sparsity within each angular image. It is already well known that two-dimensional images can be very accurately represented by only a small number of coefficients of a *sparsifying* transform, such as wavelet transforms or total-variation (TV) function applied on the image. In the case of light field images, there is another fundamental source of sparsity, that is, the angular images are very closely related to each other. Specifically, each angular image can be accurately estimated from another one using dense warping (or correspondence) fields, as shown below.

Based on the above, we use the following factorized form of the prior distribution

$$p(\mathbf{x}|\boldsymbol{\alpha}_{\mathrm{TV}}, \boldsymbol{\alpha}_c) = p(\mathbf{x}|\boldsymbol{\alpha}_{\mathrm{TV}}) \, p(\mathbf{x}|\boldsymbol{\alpha}_c) \, C(\boldsymbol{\alpha}_{\mathrm{TV}}, \boldsymbol{\alpha}_c) \qquad (10)$$

where $p(\mathbf{x}|\boldsymbol{\alpha}_{\mathrm{TV}})$ is the TV image prior employed on each angular image separately, $p(\mathbf{x}|\boldsymbol{\alpha}_c)$ is the prior that models the sparsity arising from the strong dependency between angular images and $C(\boldsymbol{\alpha}_{\mathrm{TV}}, \alpha_c)$ is a function of the unknown hyperparameters needed for the image prior model to integrate to one. In this work, we assume $C(\boldsymbol{\alpha}_{\mathrm{TV}}, \alpha_c)$ is constant.

Next we describe the specific models used for each of the prior distributions in this factorization.

1) Total Variation Image Prior: The angular images $\mathbf{x}^i$ are natural images, hence they are expected to be mostly smooth except at a number of discontinuities (e.g., spatial edges). As spatial domain image priors, we employ the total variation function which, due to its edge-preserving property, does not over-penalize discontinuities in the image while imposing smoothness [40]. Specifically, $p(\mathbf{x}|\boldsymbol{\alpha}_{\mathrm{TV}})$ is expressed as

$$p(\mathbf{x}|\boldsymbol{\alpha}_{\mathrm{TV}}) \propto \prod_{i=1}^{N} \left(\alpha_{\mathrm{TV}}^i\right)^{P/2} \exp\left[-\frac{1}{2}\alpha_{\mathrm{TV}}^i \mathrm{TV}(\mathbf{x}^i)\right] \qquad (11)$$

with

$$\mathrm{TV}(\mathbf{x}^i) = \sum_k \sqrt{(\Delta_k^h(\mathbf{x}^i))^2 + (\Delta_k^v(\mathbf{x}^i))^2} \qquad (12)$$

where Δ_k^h and Δ_k^v correspond to, respectively, horizontal and vertical first order differences, at pixel k, that is, $\Delta_k^h(\mathbf{x}^i) = (\mathbf{x}^i)_k - (\mathbf{x}^i)_{l(k)}$ and $\Delta_k^v(\mathbf{x}^i) = (\mathbf{x}^i)_k - (\mathbf{x}^i)_{a(k)}$, where $l(k)$ and $a(k)$ denote the nearest neighbors of pixel k, to the left and above, respectively.

2) Cross-Image Prior: As mentioned above, there is a high correlation between angular images in the light field image. Specifically, disregarding occlusions, each angular image $\mathbf{x}^i$ can be very closely approximated by another angular image $\mathbf{x}^j$ using the dense warping field $\mathbf{M}^{ij}$ between i and j, that is, $\mathbf{x}^i \approx \mathbf{M}^{ij}\mathbf{x}^j$. Therefore, the dependency of each angular image on another one is very strong and can be exploited while modeling $\mathbf{x}$. Based on this, we use the following cross-image prior between angular images

$$p(\mathbf{x}|\boldsymbol{\alpha}_c) \propto \exp\left(\sum_{i=1}^{N} \sum_{j \in \Omega(i)} -\frac{\alpha_c^{ij}}{2} \left\|\mathbf{x}^i - \mathbf{M}^{ij}\mathbf{x}^j\right\|_{\mathbf{O}^{ij}}^2\right) \qquad (13)$$

$$= \exp\left(\sum_{i=1}^{N} \sum_{j \in \Omega(i)} -\frac{\alpha_c^{ij}}{2} \left(\mathbf{x}^i - \mathbf{M}^{ij}\mathbf{x}^j\right)^T \mathbf{O}^{ij} \left(\mathbf{x}^i - \mathbf{M}^{ij}\mathbf{x}^j\right)\right) \qquad (14)$$

where α_c^{ij} is the precision of the registration error, and $\mathbf{O}^{ij}$ is a diagonal matrix with 0 and 1's on the diagonal to account for occlusions. In the occluded areas, the corresponding entries are set equal to zero, and the remaining entries equal to 1. This usage of the weighted norm is equivalent to the assumption that $\mathbf{O}^{ij}\mathbf{x}^i \approx \mathbf{O}^{ij}\mathbf{M}^{ij}\mathbf{x}^j$, that is, the angular image $\mathbf{x}^i$ can be closely approximated by the warped angular image $\mathbf{M}^{ij}\mathbf{x}^j$ except at the occluded areas. Notice that the occluded areas (hence matrices $\mathbf{O}^{ij}$) can easily be extracted if the warping fields $\mathbf{M}^{ij}$ are known.

In (14), $\Omega(i)$ defines a neighborhood of $\mathbf{x}^i$, which consists of the angular images with closest viewpoints to that of $\mathbf{x}^i$ (a maximum of 8 images on a rectangular grid). In other words, angular images captured by nearby regions in the aperture are treated as neighboring images. This neighborhood is imposed in (14) for several reasons. First, angular images far apart in the aperture can be less accurately related by dense warping fields due to the 3D structure of the scene and increased size of the occluded areas. Second, incorporating a cross-image prior between each pair of angular images in $\mathbf{x}$ largely increases memory requirements and therefore it is computationally not efficient during the

reconstruction phase. Finally, since $\mathbf{x}^i$ is part of at least one neighborhood defined on $\mathbf{x}$, the warping constraint is propagated to all angular images during the reconstruction algorithm.

The cross-image prior in (14) can be written in matrix-vector form as

$$p(\mathbf{x}|\boldsymbol{\alpha}_c) = z_c \exp\left(-\frac{1}{2}\mathbf{x}^T \Pi \mathbf{x}\right) \quad (15)$$

where z_c is the partition function, and Π is a sparse $NP \times NP$ matrix constructed from $N \times N$ blocks of size $P \times P$. Its explicit form is given by

$$\Pi = \begin{pmatrix} \Pi_{11} & \Pi_{12} & . & . & \Pi_{1N} \\ \Pi_{21} & \Pi_{22} & . & . & \Pi_{2N} \\ . & . & . & . & . \\ . & . & . & . & . \\ \Pi_{N1} & \Pi_{N2} & . & . & \Pi_{NN} \end{pmatrix}. \quad (16)$$

The $P \times P$ block Π_{ij} can be obtained from (14) as

$$\Pi_{ij} = \begin{cases} \sum_{s \in \Omega(i)} \alpha_c^{is} \mathbf{O}^{is} + \alpha_c^{si} \left(\mathbf{M}^{si}\right)^T \mathbf{O}^{si} \mathbf{M}^{si} & \text{if } i = j \\ -\alpha_c^{ij} \mathbf{O}^{ij} \mathbf{M}^{ij} - \alpha_c^{ji} \mathbf{O}^{ij} \left(\mathbf{M}^{ji}\right)^T & \text{if } j \neq i, \\ & \quad j \in \Omega(i) \\ 0 & \text{else} \end{cases}$$

The form of the matrix Π makes the calculation of the partition function z_c of the distribution in (15) intractable. To overcome this difficulty, we approximate the partition function by a quadratic form, and use the following as the cross-image prior

$$p(\mathbf{x}|\boldsymbol{\alpha}_c) = c \left[\prod_{i,j} \left(\alpha_c^{ij}\right)^{P/2}\right] \exp\left(-\frac{1}{2}\mathbf{x}^T \Pi \mathbf{x}\right) \quad (17)$$

with c being a constant.

It is clear that incorporating the cross-image prior requires knowledge of the dense warping fields $\mathbf{M}^{ij}$, which cannot be directly obtained from the compressive measurements. In this work, we overcome this problem by acquiring two additional images from two opposite diagonal sides of the aperture. These images exhibit full horizontal and vertical parallax, and a dense registration algorithm based on graph-cuts [41] is employed to obtain the warping field from them. Due to the uniform partitioning of the aperture, this warping field can be used to obtain approximate intermediate warping fields between all angular images. The disadvantage of this approach is that two additional exposures have to be taken with small apertures (and therefore with low SNR), and combined with the approximate calculation of the intermediate warping fields, the constraints imposed in the cross-image prior might not fully characterize the actual relations within the light field image. However, our experiments have shown that this approach provides accurate enough warping fields so that accurate reconstructions are obtained. Moreover, estimating the precision variables α_c^{ij} along with the image compensates for the inaccuracies in the warping fields during reconstruction.

An alternative method is to use $\mathbf{x}^i \approx \mathbf{x}^j$, which is similar to the approximation used in the compressive video sensing

algorithm in [42]. Although this method does not require knowledge of the warping fields, it is a very crude approximation and therefore does not provide reconstruction results comparable to the ones reported here. However, it can be used with relatively high performance in the case of very densely packed angular images, since the variation between two neighboring angular images will be very small.

It should be emphasized that the modeling in (14) is an approximation to the structure within the light field image. It implicitly assumes that the scene is Lambertian, and that the occluded areas between neighboring angular images are relatively small in size. Nevertheless, it provides a close approximation to the light-field structure (especially with small-baseline angular images as considered in this paper), and as shown in the experimental results section, it leads to a high reconstruction performance. Without such an enforcement of the internal structure of the light-field (i.e., without the use of the cross-image priors) and by only using separate image priors on the angular images, accurate reconstructions cannot be obtained. On the other hand, the role of the intra-image priors is to individually impose smoothness on the angular image estimates while preserving the sharp image features, and the advantages of employing them are demonstrated in a number of works in the literature (see, e.g., [43]).

C. Hyperpriors on the Hyperparameters

The form of the hyperprior distributions on the hyperparameters β, $\boldsymbol{\alpha}_{TV}$ and $\boldsymbol{\alpha}_c$ determines the ease of calculation of the posterior distribution $p(\mathbf{x}, \beta, \boldsymbol{\alpha}_{TV}, \boldsymbol{\alpha}_c|\mathbf{y})$. Since the distributions $p(\mathbf{y}|\mathbf{x}, \beta)$ and $p(\mathbf{x}|\boldsymbol{\alpha}_c)$ are Gaussian distributions, and we will approximate the distribution $p(\mathbf{x}|\boldsymbol{\alpha}_{TV})$ by a Gaussian distribution (shown in Section V), we chose to utilize Gamma distributions for all hyperparameters, as it is the conjugate prior for the inverse variance (precision) of the Gaussian distribution [44]. Thus, the hyperprior distributions are given by

$$p(\beta) = \Gamma(\beta|a^o, b^o) = \frac{(b^o)^{a^o}}{\Gamma(a^o)} \beta^{a^o-1} \exp\left[-b^o\beta\right] \quad (18)$$

$$p(\alpha_{TV}^i) = \Gamma(\alpha_{TV}^i|a^o, b^o), \qquad i = 1, \ldots, N \quad (19)$$

$$p(\alpha_c^{ij}) = \Gamma(\alpha_c^{ij}|a^o, b^o), \qquad i = 1, \ldots, N, \ j \in \Omega(i) \quad (20)$$

with identical shape and inverse scale parameters a^o and b^o, respectively. These parameters are set equal to small values (e.g., 10^{-5}) to make the hyperpriors vague, which makes the estimation process depend more on the observations than the prior knowledge. Note, however, that if some prior knowledge about the hyperparameters is available (for example, approximate values of the noise variances in the observations), this knowledge can easily be incorporated into the estimation procedure using appropriate values of the shape and inverse scale parameters (see, for example, [43]).

V. RECONSTRUCTION ALGORITHM

Let us denote by $\Theta = \{\beta, \boldsymbol{\alpha}_{TV}, \boldsymbol{\alpha}_c, \mathbf{x}\}$ the set of all unknowns. The Bayesian inference is based on the posterior distribution

$$p(\Theta \mid \mathbf{y}) = p(\mathbf{x}, \beta, \boldsymbol{\alpha}_{TV}, \boldsymbol{\alpha}_c|\mathbf{y}) = \frac{p(\mathbf{x}, \beta, \boldsymbol{\alpha}_{TV}, \boldsymbol{\alpha}_c, \mathbf{y})}{p(\mathbf{y})} \quad (21)$$

where $p(\beta, \boldsymbol{\alpha}_{\mathrm{TV}}, \boldsymbol{\alpha}_c, \mathbf{x}, \mathbf{y})$ is given by

$$p(\mathbf{y}, \mathbf{x}, \beta, \boldsymbol{\alpha}_{\mathrm{TV}}, \boldsymbol{\alpha}_c) = p(\mathbf{y}|\mathbf{x}, \beta)p(\mathbf{x}|\boldsymbol{\alpha}_{\mathrm{TV}}, \boldsymbol{\alpha}_c)p(\beta)p(\boldsymbol{\alpha}_{\mathrm{TV}})p(\boldsymbol{\alpha}_c). \tag{22}$$

Unfortunately, the posterior $p(\Theta \mid \mathbf{y})$ is intractable (since $p(\mathbf{y})$ is intractable), and therefore approximations are utilized. A common approximation is to represent the posterior by a delta function at its mode. Then, using $p(\mathbf{x}|\mathbf{y}, \Omega) \propto p(\Theta, \mathbf{y})$, the unknowns can be found by

$$\Theta = \arg\max_{\Theta} p(\Theta \mid \mathbf{y}) = \arg\max_{\Theta} p(\Theta, \mathbf{y})$$

Note that this formulation results in the well-known *maximum a posteriori* (MAP) estimate of Θ. Specifically, assuming uniform hyperpriors on the hyperparameters, the estimates found by this inference procedure are equivalent to the solution of the following regularized inverse problem:

$$\Theta = \arg\min_{\Theta} \left[\beta \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel^2 + \sum_{i=1}^{N} \alpha_{\mathrm{TV}}^i \mathrm{TV}(\mathbf{x}^i) \right.$$
$$\left. + \sum_{i=1}^{N} \sum_{j \in \Omega(i)} \alpha_c^{ij} \parallel \mathbf{x}^i - \mathbf{M}^{ij}\mathbf{x}^j \parallel_{\mathbf{O}^{ij}}^2 + \log z_{\boldsymbol{\alpha}} \right] \tag{23}$$

where $\mathbf{z}_{\boldsymbol{\alpha}} = \prod_{i,j} (\alpha_{\mathrm{TV}}^i)^{P/2}(\alpha_c^{ij})^{P/2}$ represents all (approximate) normalizing terms in $p(\mathbf{x}|\boldsymbol{\alpha}_{\mathrm{TV}}, \boldsymbol{\alpha}_c)$. Therefore, existing methods for TV-regularized optimization can also be employed for solving the recovery problem (see, for example, [45], [46]). However, even with the MAP approximation, the calculation of the hyperparameters is hard due to the use of the TV priors. Therefore, we resort to the majorization-minimization method proposed in [43]. We omit the details of the derivations here, and provide only the form of the updates of each unknown variable.

The estimate for the light field image $\hat{\mathbf{x}}$ can be calculated as

$$\hat{\mathbf{x}} = \Sigma_{\mathbf{x}} \beta \mathbf{A}^T \mathbf{y} \tag{24}$$
$$\Sigma_{\mathbf{x}}^{-1} = \mathrm{diag}\left(\alpha_{\mathrm{TV}}^i (\Delta^h)^T \mathbf{W}_{\mathrm{TV}}^i(\Delta^h) + \alpha_{\mathrm{TV}}^i(\Delta^v)^T \mathbf{W}_{\mathrm{TV}}^i(\Delta^v) \right)$$
$$+ \Pi + \beta \mathbf{A}^T \mathbf{A} \tag{25}$$

where the first matrix term in (25) is a $NP \times NP$ block diagonal matrix created by $P \times P$ blocks $\alpha_{\mathrm{TV}}^i (\Delta^h)^t \mathbf{W}_{\mathrm{TV}}^i(\Delta^h) + \alpha_{\mathrm{TV}}^i(\Delta^v)^t \mathbf{W}_{\mathrm{TV}}^i(\Delta^v)$. The matrices $\mathbf{W}_{\mathrm{TV}}^i$ are calculated by

$$\mathbf{W}_{\mathrm{TV}}^i = \mathrm{diag}\left(\frac{1}{\sqrt{(\mathbf{w}_{\mathrm{TV}}^i)_k}} \right), \qquad k = 1, \dots P \tag{26}$$

where

$$\left(\mathbf{w}_{\mathrm{TV}}^i \right)_k = (\Delta_k^h(\hat{\mathbf{x}}^i))^2 + (\Delta_k^v(\hat{\mathbf{x}}^i))^2. \tag{27}$$

It is clear that the vector $\mathbf{w}_{\mathrm{TV}}^i$ (and hence the matrix $\mathbf{W}_{\mathrm{TV}}^i$) represents the local spatial activity in each angular image $\mathbf{x}^i$ using its total variation. The estimates of the hyperparameters

Fig. 2. (a) Our prototype camera where an LCD array is mounted to the lens of a digital camera. (b) LCD array showing an example mask combination.

are given by

$$\beta = \frac{\frac{1}{2}NP + a^o - 1}{\frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel^2 + b^o} \tag{28}$$

$$\alpha_{\mathrm{TV}}^i = \frac{\frac{1}{2}P + a^o - 1}{\sum_k (\mathbf{w}_{\mathrm{TV}}^i)_k + b^o} \tag{29}$$

$$\alpha_c^{ij} = \frac{\frac{1}{2}P + a^o - 1}{\frac{1}{2} \parallel \mathbf{x}^i - \mathbf{M}^{ij}\mathbf{x}^j \parallel_{\mathbf{O}^{ij}}^2 + b^o}. \tag{30}$$

Finally, the algorithm iterates among estimating the light field image using (24), the spatial adaptivity vectors using (27), and the hyperparameters using (28)-(30) until convergence.

VI. PROTOTYPE LIGHT FIELD CAMERA

We have assembled a prototype of the proposed system as shown in Fig. 2(a). A binary LCD array (Electronic Assembly DOGL128S-6), shown in Fig. 2(b), is mounted to the lens of a digital camera. The LCD array consists of 128×64 pixels and we used 8×8 pixel segments as the aperture blocks. To avoid excessive vignetting, we only use the central 56×40 pixel part of the LCD array as the mask; the remaining part outside this area is covered with black carton to block light. Note that since the LCD array is binary, uniform measurement matrices cannot be realized with this mask, which require LCDs that can produce gray-scale values.

Both the LCD array and the digital camera are controlled by a computer. A specifically designed computer program successively changes the LCD image and makes an acquisition using the camera. The delay between the mask display and exposure is negligible, hence the total acquisition time approximately consists of the exposure times of each image.

Since the LCD array is not designed for this purpose, there are multiple sources of imperfections in the acquisitions[4]. For instance, the diffraction due to pixel boundaries in the LCD causes some artifacts in the acquired images. More importantly, a black pixel in the LCD array does not completely block light passing through it, which changes the effective measurement matrix. Similarly, a white pixel does not completely pass the light. Also, the pixels in the LCD array have different responses, which cause inhomogeneity within the images. To compensate for these effects, we have acquired

---

[4]Notice also that the LCD array is not placed right at the aperture plane, but in front of the lens, as the former requires extensive modification of the main lens. This also introduces some artifacts in the angular images.
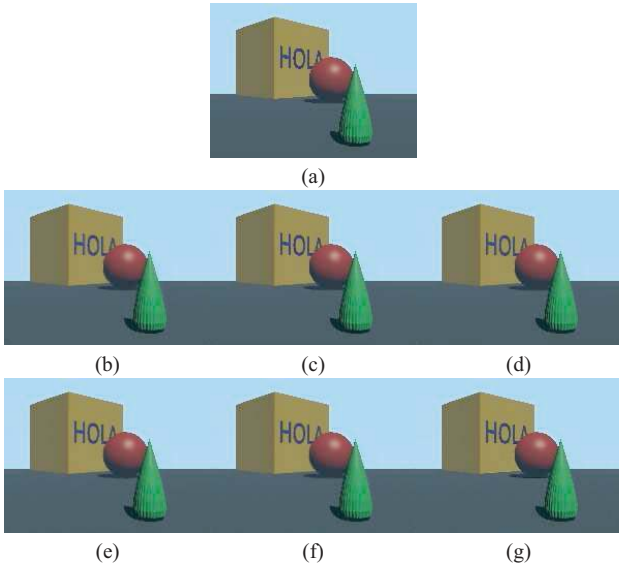
Fig. 3. Reconstruction examples with uniform matrices. (a) Original angular image. Reconstructed images from (b) and (e) 9 measurements, (c) and (f) 13 measurements, and (d) and (g) 17 measurements. (b)–(d) Low-noise case. (e)–(g) High-noise case.
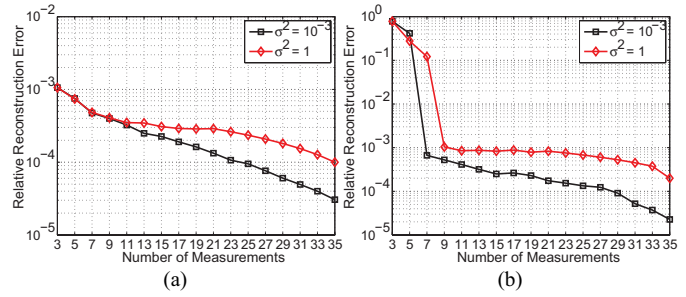

Fig. 4. Number of measurements versus relative reconstruction error (averaged over 20 runs) at two different noise levels (a) with uniform and (b) with scrambled Hadamard measurement matrices.

images of white backgrounds and color calibration boards, and used these acquisitions to approximately calculate the pixel responses. These pixel responses are then used to calculate the actual measurement mask. Although this calibration significantly reduced artifacts, this prototype system can be considerably improved with specially designed hardware. For instance, [33] recently reported that Liquid Crystal on Silicon devices are more suitable for aperture coding than LCDs. Our incoherent acquisition and reconstruction framework can be directly applied to this system as well.

VII. EXPERIMENTAL RESULTS

A. Synthetic Experiments

For synthetic experiments, a 4D light field image is constructed using the Blender software [47]. We constructed a toy 3D scene with three objects at different depths, and the camera is moved vertically and horizontally to acquire angular images that compose the 4D light field image with both horizontal and vertical parallax. One angular image from this set is shown Fig. 3(a). The light field image has a spatial resolution of 200×150 and an angular resolution of 5×7. The warping fields between the angular images are assumed to be known to test the best-case reconstruction performance. Our current (unoptimized) MATLAB implementation takes about 10 minutes on a 3GHz Core2 Duo CPU to obtain the final reconstructions.

We experimented with two different measurement matrices **A**: 1) The uniform spherical ensemble, where the entries of **A** are drawn from a uniform distribution and are between 0 and 1, and 2) scrambled Hadamard matrices, where a random subset of rows of a S-matrix [48] is chosen to generate **A**. In both cases, the expected mean of the entries in one row of **A** is 0.5, as the mean of the distribution is 0.5 in the first case and due to the property of the Hadamard matrices in the second case. Therefore, the expected amount of light

passing through the aperture in each acquisition is half of the maximum possible with both measurement matrices. Finally, we add zero-mean Gaussian noise to the measurements to obtain the final observations. We tested the reconstruction performance at two different noise levels with corresponding variances 0.001 and 1, where the intensity interval of the images is [0, 255].

We vary the number of acquired images M from 3 to 35 and apply the proposed reconstruction algorithm using the incoherent observations to obtain estimates of the original light field image. The relative reconstruction error is calculated according to $\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 / \|\mathbf{x}\|_2^2$, where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the original and estimated images, respectively.

Average reconstruction errors over 20 runs are shown in Fig. 4. Multiple remarks can be made from this figure: First, using uniform ensembles as measurement matrices generally result in lower reconstruction errors than in the case of scrambled Hadamard matrices. This is an expected result, as the uniform measurement matrices collect information from all angular images at each acquisition whereas when Hadamard matrices are used, only some of the acquired images contain information about a particular angular image. Therefore, more acquisitions are generally required to achieve the same reconstruction error.

Second, note that when the number of acquisitions is very low, e.g., 3-7, in some cases the algorithm is unable to provide successful restorations with Hadamard measurements, whereas we can always obtain some estimate of the light field image with the uniform measurements. Note, however, that although uniform measurements are clearly superior to Hadamard measurements with a low number of measurements, they achieve almost the same reconstruction performance when the number of measurements is higher than 11. This is an important result as the practical application of Hadamard matrices is much easier than employing masks with uniform measurements.

Note that the difference in the reconstruction errors between low- and high-noise cases is not significant. It is clear that the reconstruction method is very successful in reconstructing the light field image when heavy noise is present. This is especially evident in the reconstruction error at full measurement ($M = 35$); the error level is around the same order as when $M \geq 9$ in the uniform measurement case and $M \geq 11$ in the Hadamard measurement case, and the visual fidelity of the reconstructed light field remains nearly unchanged.
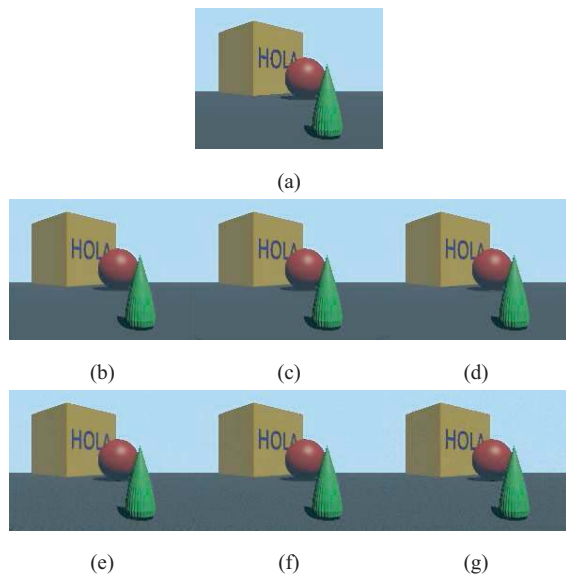
Fig. 5. Reconstruction examples with scrambled Hadamard matrices. (a) Original angular image. Reconstructed images from (b) and (e) 9 measurements, (c) and (f) 13 measurements, and (d) and (g) 17 measurements. (b)–(d) Low-noise case. (e)–(g) High-noise case.
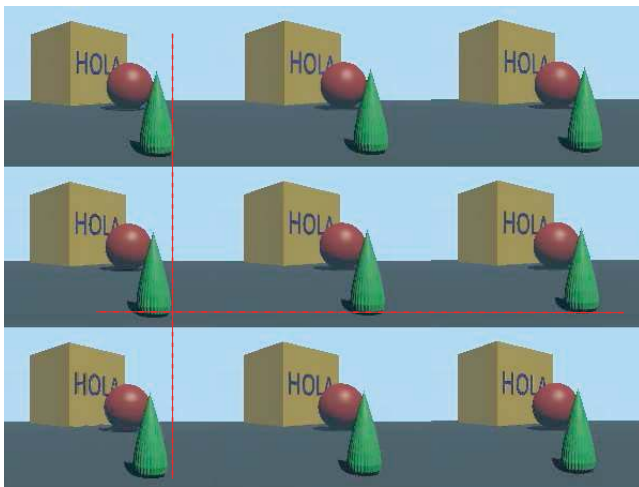
Fig. 6. Nine angular images from the light field reconstructed with 13 uniform measurements and $\sigma^2 = 10^{-3}$. The displayed images are smaller compared to the original images.
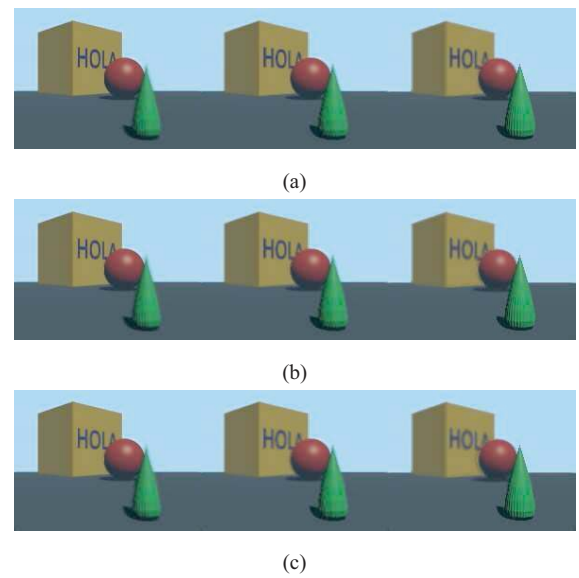
Fig. 7. Digital refocusing examples using (a) original light field image. (b) Reconstructions using uniform matrices from nine measurements. (c) Reconstructions using the scrambled Hadamard matrices from nine measurements.
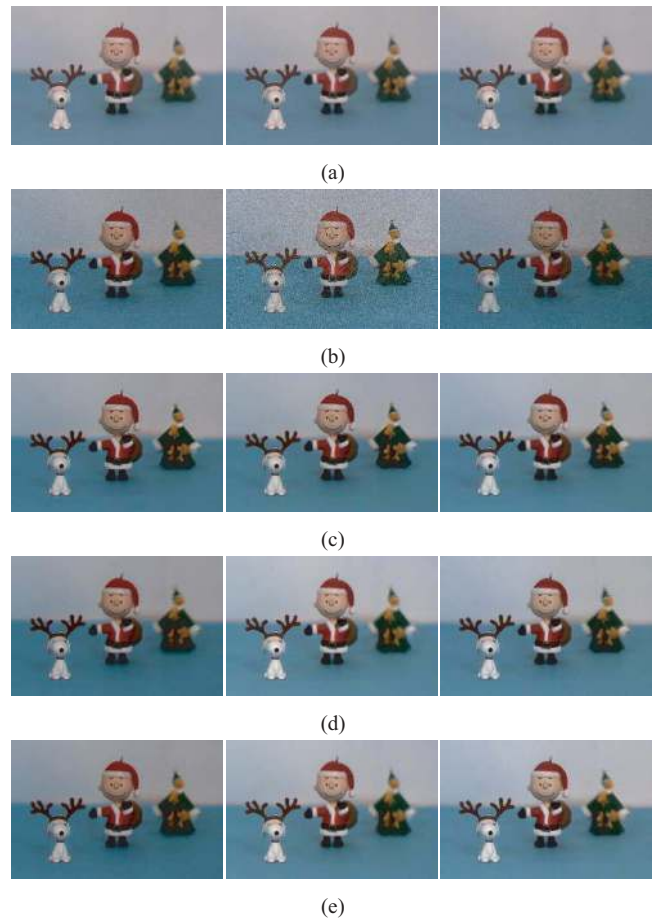
Fig. 8. Reconstruction results from a real dataset. (a) Three of the acquired images. (b) Reconstructed images using linear Hadamard inversion from 35 images. Reconstructed images using the proposed scheme from (c) 10, (d) 15, and (e) 20 acquired images.

Overall, it is clear that very accurate reconstructions can be obtained using few measurements compared to the angular dimension of the light field image. In the low-noise case, average reconstruction errors of around 6×10^{-4} and 3×10^{-4} from 9 and 15 measurements, respectively, are obtained with uniform measurement matrices. With scrambled Hadamard matrices, same error levels are achieved with about 11 and 17 measurements.

For visual quality assessment, examples of reconstructed images using 9, 13 and 17 measurements are shown in Fig. 3 for uniform and in Fig. 5 for Hadamard matrices, respectively. Note that in both cases, the reconstructed images are very close to the original angular image; the image details and structure of the scene are accurately reconstructed. The visual quality of the reconstructions can also be observed from Fig. 6, where nine angular images from the light field reconstructed from 13 measurements with uniform matrices are shown.
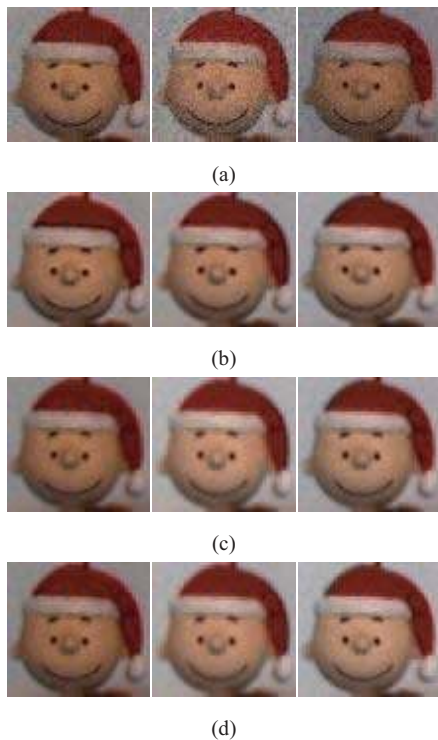
(a)

(b)

(c)

(d)

Fig. 9. Detailed parts from Fig. 8. (a) Reconstruction results using linear Hadamard inversion from 35 images. Reconstruction results using the proposed scheme from (b) 10, (c) 15, and (d) 20 acquired images.

Light field images have a number of applications in image based rendering, with typical ones being novel view synthesis and refocusing. To assess the visual quality of the reconstructed images in such a postprocessing application, we present digital refocusing results in Fig. 7. Notice that although only the reconstructed 35 images are used to obtain the refocused images, the results are of high visual quality without ghosting artifacts. Moreover, the refocused images using the reconstructions are nearly indistinguishable from the refocused images rendered using the original light field image.

In summary, it can be observed that using the proposed design the number of acquisitions can be significantly reduced (by a factor between 4 and 6). Furthermore, the reduction in the number of acquisitions is expected to be much higher with larger light field images, due to the increased level of sparsity.

B. Experiments With Real Images

Using the camera described in Section VI, we have acquired a real light field image of a representative scene. The acquired light field has angular dimensions 5×7, and each acquired image is around 10 megapixels (3888×2592). To reduce the computational load for demonstration purposes, we cropped and downsampled them to 350×230 images. We have acquired a full set of measurements (total of 35 exposures) with Hadamard measurements to compare the compressive sensing reconstruction with linear reconstruction (such as the method in [17]). The warping fields are obtained by acquiring the single-block images from the opposite ends of the mask, and using the procedure described in Section IV-B.2.

Three of the 35 acquired images are shown in Fig. 8(a). Three angular images reconstructed using linear Hadamard inversion from the full set of 35 images are shown in Fig. 8(b). The amplified noise level is clearly visible, which is not surprising since no postprocessing (such as denoising) is applied to handle the noise during the acquisition and multiplexing phase. Figures 8(c)-(e) show corresponding reconstructed angular images using the proposed scheme with 10, 15 and 20 measurements, respectively. The central parts of the images are shown in Fig. 9 for a closer inspection. Although much fewer acquisitions are used, the quality of the reconstructed images is higher than using linear reconstruction with the full dataset. It is clear that the proposed method successfully controls the trade-off between noise amplification and smoothness of the solution, thus resulting in noise-free images with sharp edges, while correcting the vignetting artifacts and nonuniform lighting to some extent. Notice that no additional postprocessing is applied to the final images to demonstrate the effectiveness of the reconstruction algorithm; the remaining illumination differences between the angular images can be corrected by employing additional postprocessing algorithms.

VIII. CONCLUSION

In this paper, we proposed a novel application of compressive sensing to a new camera design to acquire 4D light field images. We have shown that incoherent measurements of angular images can be collected by using a randomly coded mask placed at the aperture of a traditional camera. These measurements are then used to reconstruct the original light field image. We developed a reconstruction algorithm which exploits the high degree of information redundancy (and hence, sparsity) inherent in the light field images, and have shown that the complete light field image can be reconstructed using only a few acquisitions. Moreover, the captured images have high signal-to-noise ratios due to small amount of loss of light. The proposed design provides high spatial and angular resolution light field images, and does not suffer from limitations of many existing light field imaging systems. Finally, the proposed design can be implemented by simple modifications of traditional cameras. Experimental results with both synthetic and real image sets show the effectiveness and potential of this approach for light field acquisition.

The proposed design, although powerful in terms of providing both high spatial- and angular-resolution, also has several limitations. Most importantly, it requires the scene and the camera to be static as a number of acquisitions have to be made. Any object or camera motion will necessarily introduce significant artifacts in the reconstructed images. In addition, our current implementation of the reconstruction method requires simultaneous processing of all angular images and the observations. Although we observed that the convergence is generally very fast, this processing might lead to high computational load if the size of the light field is large. Although not explored in this paper, this problem can potentially be addressed by processing images in patches and by parallel processing.

REFERENCES

[1] F. Durand and R. Szeliski, "Guest editors' introduction: Computational photography," *IEEE Comput. Graph. Appl.*, vol. 27, no. 2, pp. 21–22, Mar.–Apr. 2007.

[2] S. D. Babacan, R. Ansorge, M. Luessi, R. Molina, and A. K. Katsaggelos, "Compressive sensing of light fields," in *Proc. IEEE Int. Conf. Image Process.*, Cairo, Egypt, Jul. 2009, pp. 2337–2340.

[3] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[4] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[5] R. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[6] E. Candès, "Compressive sampling," in *Proc. Int. Congr. Math.*, Madrid, Spain, 2006, pp. 1433–1452.

[7] F. Ives, "Parallax stereogram and process of making same," U.S. Patent 3 725 567, Apr. 14, 1903.

[8] G. Lippmann, "Epreuves reversible donnant la sensation du relief," *J. Phys.*, vol. 7, no. 4, pp. 821–825, Nov. 1908.

[9] T. Adelson and J. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 99–106, Feb. 1992.

[10] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Tech.*, 1996, pp. 31–42.

[11] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Tech.*, 1996, pp. 43–54.

[12] R. Ng, M. Levoy, M. Brèdif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Dept. Comput. Sci., Stanford Univ., Stanford, CA, Tech. Rep. CTSR 2005-02, 2005.

[13] T. Georgiev, K. C. Zheng, B. Curless, D. Salesin, S. Nayar, and C. Intwala, "Spatio-angular resolution tradeoffs in integral photography," in *Proc. Eurograph. Symp. Render.*, Jun. 2006, pp. 263–272.

[14] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, 2005.

[15] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 69-1–69-12, Jul. 2007.

[16] T. Georgiev, C. Intwala, S. D. Babacan, and A. Lumsdaine, "Unified frequency domain analysis of lightfield cameras," in *Proc. Eur. Conf. Comput. Vis.*, Marseille, France, Dec. 2008, pp. 1–15.

[17] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen, "Programmable aperture photography: Multiplexed light field acquisition," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–10, Aug. 2008.

[18] A. Ashok and M. Neifeld, "Compressive light field imaging," *Proc. SPIE 3-D Visuali. Process. II*, vol. 7690, p. 76900Q, Apr. 2010.

[19] E. E. Fenimore and T. M. Cannon, "Coded aperture imaging with uniformly redundant arrays," *Appl. Opt.*, vol. 17, no. 3, pp. 337–347, 1978.

[20] S. R. Gottesman and E. E. Fenimore, "New family of binary arrays for coded aperture imaging," *Appl. Opt.*, vol. 28, no. 20, pp. 4344–4352, 1989.

[21] R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: Motion deblurring using fluttered shutter," in *Proc. SIGGRAPH*, 2006, pp. 795–804.

[22] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," in *Proc. ACM Graph. SIGGRAPH Conf.*, 2007, p. 70.

[23] A. Zomet and S. K. Nayar, "Lensless imaging with a controllable aperture," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 339–346.

[24] S. K. Nayar and V. Branzoi, "Adaptive dynamic range imaging: Optical control of pixel exposures over space and time," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1168–1175.

[25] A. Mohan, X. Huang, J. Tumblin, and R. Raskar, "Sensing increased image resolution using aperture masks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[26] H. Farid and E. P. Simoncelli, "Range estimation by optical differentiation," *J. Opt. Soc. Amer. A*, vol. 15, no. 7, pp. 1777–1786, 1998.

[27] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Exp.*, vol. 15, no. 21, pp. 14013–14027, 2007.

[28] P. Sen and S. Darabi, "Compressive dual photography," *Comput. Graph. Forum*, vol. 28, no. 2, pp. 609–618, 2009.

[29] J. Gu, S. K. Nayar, E. Grinspun, P. N. Belhumeur, and R. Ramamoorthi, "Compressive structured light for recovering inhomogeneous participating media," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2008, pp. 1–14.

[30] R. Marcia and R. Willett, "Compressive coded aperture superresolution image reconstruction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2008, pp. 833–836.

[31] R. F. Marcia, Z. T. Harmany, and R. M. Willett, "Compressive coded aperture imaging" *Proc. SPIE, Comput. Imag.*, vol. 7246, no. 1, pp. 72460G-1–72460G-13, Feb. 2009.

[32] W. T. Cathey and E. R. Dowski, "New paradigm for imaging systems," *Appl. Opt.*, vol. 41, no. 29, pp. 6080–6092, 2002.

[33] H. Nagahara, C. Zhou, T. Watanabe, H. Ishiguro, and S. Nayar, "Programmable aperture camera using LCoS," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 337–350.

[34] C.-K. Liang, Y.-C. Shih, and H. Chen, "Light field analysis for modeling image formation," *IEEE Trans. Image Process.*, vol. 20, no. 2, pp. 446–460, Feb. 2011.

[35] Z. Zhang and M. Levoy, "Wigner distributions and how they relate to the light field," in *Proc. Int. Conf. Comput. Photography*, Apr. 2009, pp. 1–10.

[36] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[37] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[38] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.

[39] L. Gan, T. Do, and T. Tran, "Fast compressive imaging using scrambled block Hadamard ensemble," in *Proc. EUSIPCO*, Lausanne, Switzerland, Aug. 2008, pp. 1–5.

[40] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, nos. 1–4, pp. 259–268, 1992.

[41] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[42] R. Marcia and R. Willet, "Compressive coded aperture video reconstruction," in *Proc. EUSIPCO*, Lausanne, Switzerland, Aug. 2008, pp. 1–5.

[43] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Parameter estimation in TV image restoration using variational distribution approximation," *IEEE Trans. Image Process.*, vol. 17, no. 3, pp. 326–339, Mar. 2008.

[44] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag, 1985, chs. 3 and 4.

[45] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," in *Handbook of Mathematical Models in Computer Vision*, N. Paragios, Y. Chen, and O. Faugeras, Eds. New York: Springer-Verlag, 2005.

[46] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty, "An efficient algorithm for compressed MR imaging using total variation and wavelets," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[47] *Blender Software* [Online]. Available: http://www.blender.org/

[48] M. Harwit and N. J. A. Sloane, *Hadamard Transform Optics*. New York: Academic, 1979.

S. Derin Babacan (M'10) received the B.Sc. degree from the Electrical and Electronics Department, Bogazici University, Istanbul, Turkey, in 2004, and the M.Sc. and Ph.D. degrees from the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, in 2006 and 2009, respectively.

He is currently a Beckman Post-Doctoral Fellow with the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana. His current research interests include inverse problems in image processing, computer vision, and computational photography.

Dr. Babacan was a recipient of the IEEE International Conference on Image Processing Paper Award in 2007.

Reto Ansorge was born in 1979. He received the Dipl.Ing. FH. degree in elektrotechnik from the HSR Hochschule für Technik Rapperswil, Rapperswil, Switzerland, in 2005, and the M.Sc. degree from the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, in 2010.

He was with HSR Medialab, Raperswil, as a Research Engineer from 2005 to 2007. Since 2009, he has been with Varian Medical Systems, Baden, Switzerland, where he is currently researching kV and MV image acquisition projects.

Martin Luessi (S'04–M'12) received the Ing. FH. degree from the HSR Hochschule für Technik Rapperswil, Rapperswil, Switzerland, in 2006, and the M.S. and Ph.D. degrees from the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, in 2007 and 2011, respectively, all in electrical engineering.

He joined the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, MA, in 2011, as a Research Fellow. He currently holds an appointment as a Research Fellow with Harvard Medical School, Boston. His current research interests include signal processing methods for neuroimaging, particularly methods for magnetoencephalography, Bayesian modeling and inferences, numerical optimization, sparse signal processing, and machine learning.

Pablo Ruiz Matarán (S'11) received the M.S. degree in mathematics and the Master's degree in multimedia technologies from the University of Granada, Granada, Spain, in 2008 and 2009, respectively, where he is currently pursuing the Ph.D. degree with the Department of Computer Science and Artificial Intelligence, Visual Information Processing Group.

He is involved in research with the Spanish research program Consolider Ingenio 2010: Multimodal Interaction in Pattern Recognition and Computer Vision. His current research interests include super-resolution and classification of multispectral satellite images, video retrieval from video databases, and fusion of visible-infrared images.

Rafael Molina (M'88) was born in 1957. He received the M.S. Degree in mathematics (statistics) and the Ph.D. degree in optimal design in linear models in 1979 and 1983, respectively.

He has been a Professor of computer science and artificial intelligence with the University of Granada, Granada, Spain, since 2000. His current research interests include image restoration and its applications to astronomy and medicine, parameter estimation in image restoration, super resolution of images and videos, and blind deconvolution.

Aggelos K. Katsaggelos (F'98) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1979, and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1981 and 1985, respectively.

He has been with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, since 1985, where he is currently a Professor and the AT&T Chair. He was the Ameritech Chair of Information Technology from 1997 to 2003. He is the Director of the Motorola Center for Seamless Communications, a member of the Academic Staff, NorthShore University Health System, an Affiliated Faculty Member with the Department of Linguistics, Northwestern University, and he has an appointment with the Argonne National Laboratory, Lemont, IL. He has authored or co-authored many papers in the areas of multimedia signal processing and communications, including more than 180 journal papers, 400 conference papers, and 40 book chapters. He holds 19 international patents. He is the co-author of *Rate-Distortion Based Video Compression* (Kluwer, 1997), *Super-Resolution for Images and Video* (Claypool, 2007), and *Joint Source-Channel Video Transmission* (Claypool, 2007).

Prof. Katsaggelos was the Editor-in-Chief of the *IEEE Signal Processing Magazine* from 1997 to 2002, a BOG Member of the IEEE Signal Processing Society from 1999 to 2001, and a member of the Publication Board of the IEEE Proceedings from 2003 to 2007. He was a fellow of the SPIE in 2009. He was a recipient of the IEEE Third Millennium Medal in 2000, the IEEE Signal Processing Society Meritorious Service Award in 2001, the IEEE Signal Processing Society Technical Achievement Award in 2010, the IEEE Signal Processing Society Best Paper Award in 2001, the IEEE International Conference on Multimedia and Expo Paper Award in 2006, the IEEE International Conference on Image Processing Paper Award in 2007, and the International Student Paper of the Year Award in 2009. He was a Distinguished Lecturer of the IEEE Signal Processing Society from 2007 to 2008.