LIGHT FIELD COMPRESSIVE SENSING IN CAMERA ARRAYS

Mahdad Hosseini Kamal, Mohammad Golbabaee and Pierre Vandergheynst

Signal Processing Laboratory (LTS2), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
E-mail:{mahdad.hosseinikamal, mohammad.golbabaei, pierre.vandergheynst}@epfl.ch

ABSTRACT

This paper presents a novel approach to capture light field in camera arrays based on the compressive sensing framework. Light fields are captured by a linear array of cameras with overlapping field of view. In this work, we design a redundant dictionary to exploit cross-cameras correlated structures to sparsely represent cameras image. Our main contributions are threefold. First, we exploit the correlations between the set of views by making use of a specially designed redundant dictionary. We show experimentally that the projection of complex scenes onto this dictionary yields very sparse coefficients. Second, we propose an efficient compressive encoding scheme based on the random convolution framework [1]. Finally, we develop a joint sparse recovery algorithm for decoding the compressed measurements and show a marked improvement over independent decoding of CS measurements.

Index Terms— Redundant Dictionary, Compressive Sensing, Light Fields, l_1-minimization.

1. INTRODUCTION

The growing flood of data, particularly due to the emergence of image processing systems with multiple visual sensors, causes an increasing need to quickly process large data sets in compressed domain. Moreover, advances in computational photography provides novel methods to acquire and process images. Newly introduced light field cameras are one of the widely used class of computational cameras. The light field cameras capture the most complete representation of the scene, the so-called plenoptic function [2]. The plenoptic function is a 5D function that models the amount of light rays a perfect observer records at any position, direction, time and wavelength in free space. A number of light field cameras have been designed to capture a subset of the plenoptic function. By capturing a set of images in the space, we can reconstruct samples of the plenoptic function, which later can be used for depth calculation or other applications like image based rendering.

The most common plenoptic cameras are those using 1D/2D camera arrays like multi-camera arrays [3]. However, capturing the light fields leads to a large amount of data, which reveals the importance of adapting an intelligent acquisition method that relies on the properties of the camera networks. We consider a plenoptic camera, which consists of array of k equally spaced cameras, to capture the light rays coming from a scene like in Fig. 1. In order to represent camera outputs, we stack all images to have an image volume $\mathcal{X} \in \mathbb{R}^{i \times j \times k}$ as shown in Fig. 2(a) in which each slice of size $i \times j$ corresponds to an image observed by the corresponding camera.

This work is funded by the Swiss National Science Foundation under grant number 200021-125651 and the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL.

Part of this work was submitted for publication to ICASSP 2012.

Fig. 1: Original scene.

In order to tackle the large amount of data produced by plenoptic cameras, we should consider a compression method. Compressive Sensing is a popular compression method that has superiority over traditional schemes because of low complexity in the encoder and universality with respect to scene model. Compressive sensing dictates to recover a signal $x \in \mathbb{R}^n$ from many fewer measurements $m \ll n$ than the traditional methods, provided that the signal is sparse or compressible in some basis $\mathbf{\Phi}$. The compressive sensing measurements are formed by taken inner product between the signal and a random measurement matrix. The measurements can be expressed as

$$y = \mathcal{A}x, \tag{1}$$

where $y \in \mathbb{R}^m$ is the measurements vector and $\mathcal{A} \in \mathbb{R}^{m \times n}$ is the measurement matrix. An approach to recover x from y is l_1 minimization in which we solve the following convex problem:

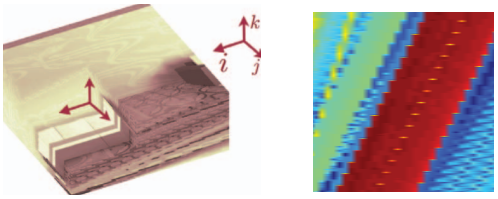
$$\underset{\alpha \in \mathbb{R}^n}{\operatorname{argmin}} \|\alpha\|_1 \quad \text{subject to} \quad y = \mathcal{A}\mathbf{\Phi}\alpha \tag{2}$$

and x is obtained from $x = \mathbf{\Phi}\alpha$. In short, the above algorithm looks for the set of transform coefficients α such that the measurements from the corresponding signal $\mathbf{\Phi}\alpha$ match the measurements y. To insure a successful recovery for x, the measurement matrix $\mathcal{A}$ should satisfy the uniform uncertainty principle [4, 5, 6]. More precisely, the measurement matrix should have a small restricted isometry constant δ. The restricted isometry constant for a S-sparse signal x is

$$(1 - \delta_S)\|x\|_2^2 \leq \|\mathcal{A}x\|_2^2 \leq (1 + \delta_S)\|x\|_2^2. \tag{3}$$

The simplest construction uses random sensing matrices with entries generated independently according to a subgaussian distribution, like independent and identically distributed (i.i.d) Gaussian or Bernoulli/Rademacher (random ± 1). The compressive sensing theory for this type of measurement matrices and a S-sparse signal in $\mathbf{\Phi}$ implies that with $m \geq O(S \log n/S)$ measurements we can reconstruct the signal.

In this paper, we present a novel compressive acquisition scenario



(a) (b)

Fig. 2: (a) Image volume. (b) (i, k)-plane slice of the image volume.

for the light field images. We design a redundant dictionary based on corss-correlations of cameras image to take advantage from the local and non-local structures in the camera array.

2. CAMERA ARRAY ACQUISITION SCHEME

The large amount of data and practical limitations in camera array highlight the importance of employing a computationally tractable measurement system. To have a feasible measurement matrix, we use the Random Convolution strategy explained in the work of J. Romberg [1] for each camera. A different physical realization of this sensing matrix is also discussed in [7]. In short, the method subsamples m random values of the signal x circularly convolved with a random filter. It is proved that a S-sparse signal x is recovered by $m \geq O(S \log n/\delta)$ $\delta \in [0, 1]$ measurements.

We acquire the measurements on camera p by $y_p = \mathcal{A}_p x_p$, where $y_p \in \mathbb{R}^m$ represents the measurement vector, $\mathcal{A}_p \in \mathbb{R}^{m \times n}$ denotes the random convolution sensing matrix on the camera, and $x_p \in \mathbb{R}^n$ with $n = i \cdot j$ is the vectorized representation of the corresponding camera image. The signal and measurement ensemble of the camera array are represented as

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathcal{A}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathcal{A}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathcal{A}_k \end{bmatrix}, \tag{4}$$

where $X \in \mathbb{R}^{kn}, Y \in \mathbb{R}^{km}$, $\mathbf{A} \in \mathbb{R}^{km \times kn}$ and $\mathbf{0}$ is zero matrix with appropriate size. The measurement vector ensemble is summarized as

$$Y = \mathbf{A}X. \tag{5}$$

3. RECOVERY SCHEME

Cameras image can be recovered by two different approaches. In the first compression method, we reconstruct each camera image separately. Cameras are placed in such a way that they have a large overlapping field-of-view. Therefore, the captured images from a camera to the others are highly correlated. The separate recovery approach can benefit from image structures in each camera but it does not consider any correlation among cameras in the camera network. The 2D wavelet transform is the distinguished sparsity domain for cameras image. Since in the acquisition and reconstruction phase cameras do not collaborate, consequently, the recovery algorithm for each camera p is summarized as

$$\underset{u_p \in \mathbb{R}^n}{\operatorname{argmin}} \|u_p\|_1 \quad \text{subject to} \quad \|y_p - \mathcal{A}_p \mathbf{\Phi} u_p\|_2 \leq \epsilon_p, \tag{6}$$

where $\mathbf{\Phi}$ is 2D wavelet transform matrix and ϵ_p is the measurement noise. The corresponding camera image is then recovered by $x_p = \mathbf{\Phi} u_p$.

In the second approach, we are going to exploit both intra- and inter-camera correlated structures in images. The framework leads to exploit the amongst camera correlations to jointly recover the images from the measurement vector ensemble and generalizes the notion of a sparse signal in some basis to the concept of signal ensemble that are jointly sparse in some domain. In this notion, each signal itself is sparse in a basis, so we could benefit from the compressive sensing theory to encode and decode each signal separately. However, this approach relies on joint sparsity [8], which is stronger than the aggregated sparsity of individual signals. As a result, the joint recovery strategy leads to a reduction in the number of required measurements. The only challenge for this framework is to find a sparsity domain for the signal ensemble X.

By the emergence of redundant dictionaries in compressive sensing, we can hope for such a dictionary in which the signal ensemble is sparsely represented. A well-designed dictionary can benefit from the data regularity in the network to sparsely represent the signal ensemble. The combination of such a dictionary with our joint recovery scheme leads to consider both local and non-local structures. Therefore, the number of required measurements will be decreased.

4. COMPRESSIVE SENSING WITH REDUNDANT DICTIONARIES

The emerging framework of compressive sensing assures that a signal can be accurately recovered from much smaller number of measurements than required by traditional methods [9]. As we explained, the compressive sensing technique holds for signals that are sparse either in their standard coordinate or in any orthogonal basis, thus it is important to find a set of basis functions that can best represent structures in signals. Although bases such as Fourier and wavelet can provide a good representation of signals, they are generic and not specific enough to very restrictive class of signals. An alternative signal representation is to consider an overcomplete dictionary $\mathcal{D} \in \mathbb{R}^{n \times d}$ with $d > n$.

In an overcomplete dictionary the signal decomposition is not unique, but this allows to finely adapt the dictionary to the expected signal structures, i.e. choosing among many representations the one that has the sparsest possible representation of the signal.

It has been shown in [10, 11] that the compressive sensing techniques for orthogonal bases can be extended to signals that are not sparse in bases but rather in redundant dictionaries. Given a suitable sensing matrix the procedure identifies the sparsest coefficient sequence θ of the signal x in the redundant dictionary $\mathcal{D}$, i.e. $x = \mathcal{D}\theta$. An approach to recover x is by solving the following convex problem:

$$\underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \|\theta\|_1 \quad \text{subject to} \quad \|y - \mathcal{A}\mathcal{D}\theta\|_2 \leq \epsilon \tag{7}$$

4.1. Dictionary Design

The geometric features of a signal are the heart of dictionary design. It is known that wavelets give the opportunity to compress piecewise smooth images. For example, a stair-like image similar to Fig. 3(a) can be potentially compressed by 2D wavelet transform, but indeed

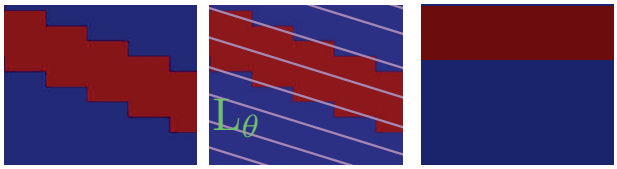


Fig. 3: The reordering example in stair-like images. (a) Original stair-like image. (b) Parallel reordering lines L_η to capture regularity along direction η. (c) Reordered 2D stair-like image along L_η.

the regularity along the directions motivates the design of a more intelligent domain to exploit these structures in the image. Our goal is to include anisotropic redundancy that cannot be captured by a simple 2D wavelet transform and use a dictionary to sparsely represent the image.

A piecewise constant function is sparsely represented in the 1D wavelet domain. Therefore, for stair-like images it would be best to consider the reordered version of the image grid to have a piecewise-constant-like images [12].

The reordering process is described by selecting a direction η, which is as parallel as possible to the real geometry of the curve. As it is shown on Fig. 3(b), we select grid points L_η and reorder the image grid according to the indices of image samples on these lines Fig. 3(c). Afterwards, we make a piecewise smooth 1D discrete function f from the reconstructed image, which can be sparsely represented using 1D wavelet domain.

If we compare the coefficients of the stair-like image in the simple 2D wavelet and reordered 1D wavelet domains, we can see a great improvement induced by our scheme. This also reveals the importance of considering image geometry in designing a dictionary. Fig. 4 contrasts the representations in 1D and 2D wavelet transform. Selecting a proper direction η for reordering lines has a direct effect on sparsity of the represented signal. The 1D wavelet transform itself provides an efficient way to distinguish the appropriate direction. An inadmissible direction for the reordering process increases the number of 1D wavelet coefficients as demonstrated in Fig. 5. The best η is the one that leads to the sparsest representation of the signal.

In the case of stair-like images with different directions, we do not have a preferential orientation. Thus we cannot represent a sparse image with just one direction, but we can benefit from a redundant dictionary, which consists of the concatenation of several reordered wavelet transform $\mathbf{\Phi}^r$ with different directions η. The union of bases redundant dictionary $\mathbf{\Psi} = \left[\mathbf{\Phi}_1^r, \mathbf{\Phi}_2^r, \cdots, \mathbf{\Phi}_\gamma^r\right]$ is destined to profit from different reordering directions. Therefore, it will exploit the geometry induced by the natural correlations within light field images. We should consider that given too many reordering directions will not result in a unique sparsest representation, since as the correlation between dictionary atoms is increased the process cannot decide which atom to choose. Moreover, choosing too many directions leads to a huge dictionary which is not efficient.

In the camera array scenario, as shown in Fig. 2, the image volume $\mathcal{X}$ would have stair-like structure along (i,k)-plane. Thus, once the dictionary $\mathbf{\Psi}$ has enough angular resolution (number of different directions γ), $\mathcal{X}$ can be efficiently represented by few sparse coefficients along (i,k)-plane. In addition, a suitable 1D wavelet transform can be applied to sparsify $\mathcal{X}$ along the remaining dimension j. This comes from the fact that natural 2D images typically

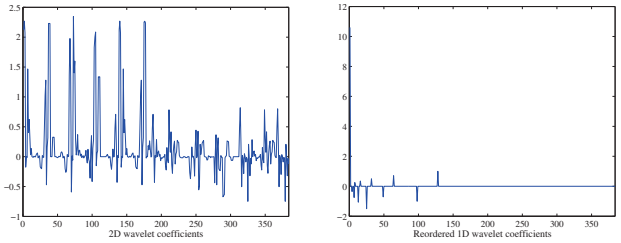


Fig. 4: Left: 2D wavelet coefficients of a stair-like images. Right: 1D wavelet coefficients of the reordered 1D discrete function.

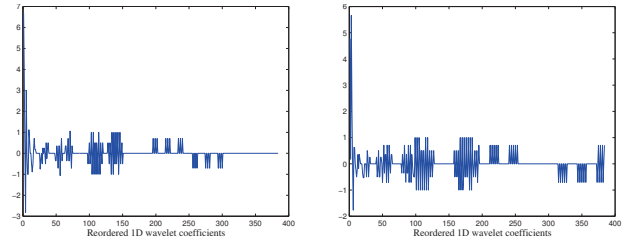


Fig. 5: Influence of choosing inappropriate directions on 1D wavelet coefficients on the reordered 1D discrete function.

have piecewise smooth variations along both dimensions i and j. To achieve an efficient representation, we reshape X into a matrix $\widehat{\mathbf{X}} \in \mathbb{R}^{ik\times j}$ whose columns contain the information of (i,k)-planes. Following the discussion above, there exists a *sparse* matrix of coefficients $\mathbf{\Theta} \in \mathbb{R}^{\gamma ik\times j}$ such that $\widehat{\mathbf{X}} = \mathbf{\Psi\Theta\Gamma}^T$ where $\mathbf{\Psi} \in \mathbb{R}^{ik\times\gamma ik}$ is the previously defined dictionary transform along (i,k)-plane and $\mathbf{\Gamma} \in \mathbb{R}^{j\times j}$ denotes the 1D wavelet basis along j dimension. Thus, if we rewrite $\widehat{\mathbf{X}}$ and $\mathbf{\Theta}$ matrices in vectorial format, we will have

$$\widehat{X}_{vec} = \mathbf{\Omega}\Theta_{vec}, \tag{8}$$

where $\mathbf{\Omega} \in \mathbb{R}^{nk\times\gamma nk}$ is the dictionary that is applied to encode the whole image volume into a sparse vector Θ_{vec} and its 3D dimensions. Note that a simple calculation reveals that $\mathbf{\Omega} = \mathbf{\Psi}\otimes\mathbf{\Gamma}$, where $\otimes$ denotes the Kronecker product between two matrices.

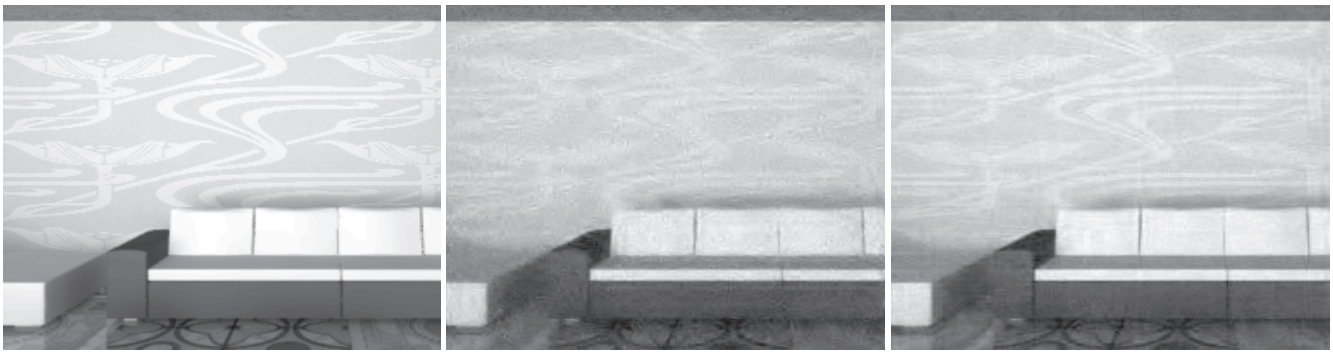
Once we have designed the proper dictionary to efficiently encode the image volume into few sparse coefficients, the following convex problem can be applied to reconstruct X from the compressive measurements,

$$\underset{\Theta_{vec}\in\mathbf{R}^{\gamma nk}}{\operatorname{argmin}} \|\Theta_{vec}\|_1 \quad \text{subject to} \quad \|Y - \widehat{\mathbf{A}}\mathbf{\Omega}\Theta_{vec}\|_2 \leq \epsilon. \tag{9}$$

Here $\widehat{\mathbf{A}}$ contains the same elements as $\mathbf{A}$, and is reshaped with respect to $\widehat{X}_{vec} = \mathbf{\Omega}\Theta_{vec}$ so that $\widehat{\mathbf{A}}\widehat{X}_{vec} = \mathbf{A}X$. Not that this optimization can be solved iteratively using Douglas-Rachford splitting method [13] and it basically consists of alternating between a shrinkage operator (soft thresholding) and projection onto the convex set $\|Y - \widehat{\mathbf{A}}\mathbf{\Omega}\Theta_{vec}\|_2 \leq \epsilon$, until converging to the solution.

5. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of our acquisition model, we generated a synthetic scene shown in Fig. 1. We captured the scene by



(a) (b) (c)

Fig. 6: Reconstruction of a camera image with separate recovery and our joint recovery scheme, by randomly sampling of each camera by 25% of its image size. The results reveal the superiority of our scheme by about $3dB$.(a) Original image. (b) Reconstruction with separate recovery scheme. (c) Reconstruction with joint recovery scheme.

$k = 40$ equally distanced cameras with resolution of 256×256. Following our aforementioned acquisition scheme, we used the random convolution measurement matrix to randomly sample each camera image by 25% of the camera original image size, i.e. $m = 0.25n$. For the redundant dictionary, we take 3 different reordering directions to capture cameras image regularities.

In order to further evaluate our joint recovery scheme, we reconstruct each camera image with two different recovery algorithms. First, we *separately* reconstruct each camera image by solving (6), which only benefits from the sparse representation of the (i, j)-planes in a 2D wavelet basis $\mathbf{\Phi}$. Therefore, we do not incorporate any intra-camera correlated structures. Second, we apply our *joint recovery* algorithm (9) in order to benefit from both inter-/intra-camera structures.

Fig. 6 compares the recovery performance of both algorithms on one camera to the original image. One of the images is reconstructed separately without benefiting from any cross-camera correlation, while the other one is reconstructed by our joint recovery algorithm in order to exploit cross-camera correlations. As we can see, with the same number of measurements, our joint recovery scheme looks more similar to the original image. In average the joint recovery scheme overtakes the separate recovery approach by about $3dB$. This highlights the role of our designed redundant dictionary, which results in exploiting correlated structures in the camera array.

6. CONCLUSION

This paper represents a novel approach to capture the light fields in a camera array based on sparse representations in redundant dictionaries. We developed a reconstruction algorithm which exploits the high degree of correlations in camera network and have shown that the complete light field image can be reconstructed using only few measurements. The proposed algorithm relies on the random convolutions scheme and can be implemented on existing hardware. Finally, simulated experiments demonstrated the potential interest of our proposed scheme.

7. REFERENCES

[1] J. Romberg, "Compressive Sensing by Random Convolution," *SIAM SIIMS*, 2009.

[2] Edward H. Adelson and James R. Bergen, "The Plenoptic Function and the Elements of Early Vision," in *Comput. Mod. Vis. Proc.*, 1991.

[3] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antnez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High Performance Imaging using Large Camera Arrays," *ACM Trans. Graph.*, 2005.

[4] D. Donoho, "Compressed Sensing," *IEEE Trans. Inform. Theory*, 2006.

[5] E. Candès, J. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," *Comm. Pure Appl. Math.*, 2005.

[6] E. Candès and J. Romberg and T. Tao, "Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?," *IEEE Trans. Inform. Theory*, 2004.

[7] L. Jacques, P. Vandergheynst, A. Bibet, V. Majidzadeh, A. Schmid, and Y. Leblebici, "CMOS Compressed Imaging by Random Convolution," *ICASSP*, 2009.

[8] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed Compressive Sensing," *CoRR*, 2009.

[9] R. G. Baraniuk, "Compressive Sensing," *IEEE Sig. Proc. Mag.*, 2007.

[10] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed Sensing and Redundant Dictionaries," *IEEE Trans. Inform. Theory*, 2008.

[11] E. Candès and C. Eldar and D. Needell, "Compressed Sensing with Coherent and Redundant Dictionaries," *Appl. Comput. Harmon. Anal.*, 2011.

[12] G. Peyré and S. Mallat, "Surface Compression with Geometric Bandelets," *ACM SIGGRAPH*, 2005.

[13] P. L. Combettes and J. C. Pesquet, "Proximal Splitting Methods in Signal Processing," *arXiv/0912.3522*, 2009.