

Compressive light field spectral imaging in a single-sensor device by using coded apertures

M. Marquez¹, N. Diaz², J. Bacca³, S. Pertuz², H. Arguello³

¹Universidad Industrial de Santander, Department of physics, Bucaramanga 680002, Colombia

²Universidad Industrial de Santander, Department of Electrical Engineering, Bucaramanga 680002, Colombia

³Universidad Industrial de Santander, Department of Systems and Informatics Engineering, Bucaramanga 680002, Colombia

henarfu@uis.edu.co

Abstract: This work introduces a light field architecture aimed at simultaneously performing multi-spectral compressive imaging and light field acquisition in a single-sensor device. The proposed architecture exploits a microlens array combined with coded aperture patterns.

1. Introduction

Light field photography allows sampling the first 5 dimensions (space and orientation) of the plenoptic function $f(x, y, z, \gamma, \theta, \lambda, t)$, where the 7-dimensional field are represented by space (x, y, z) , orientation (γ, θ) , wavelength λ , and times t . In the literature, single-sensor light field architectures can be grouped into two categories [1]. In the first one, mask-based designs are exploited for compressive light field sensing [2–4]. This approach has the advantage of yielding high-resolution light field photographs at the cost of reduced light transmission and higher computational complexity. In the second category, an array of lenslets is placed in front of the sensor within the optical path of a monocular camera. This approach has the advantage of an improved light gathering and extended depth of field at the cost of a reduced spatial resolution. This paper introduces an architecture aimed at simultaneously performing multi-spectral compressive imaging and light field acquisition in a single-sensor device. The proposed architecture integrates multi-spectral imaging, by means of the coded-aperture approach, with light field sampling by placing a lenslet array in front of the imaging sensor. The computational design parameters associated with compressed sensing of multi-spectral images in the presence of the lenslet array are studied. In particular, several sparse representation bases are studied in order to determine the one yielding the best performance. Subsequently, different data structures for arranging the multi-dimensional plenoptic function are proposed and assessed in order to improve computational efficiency and reconstruction accuracy. Experiments are performed using synthetic multi-spectral light field and results demonstrate that the proposed architecture allows the reconstruction of the plenoptic function with a PSNR up to 33.75 dB. In the literature, the more closely related approaches are aimed either at sampling the first 5-dimensions of the plenoptic function $(x, y, z, \gamma, \theta)$, such as in light field imaging, or capturing spectral information (x, y, λ) , such as in multi-spectral imaging. The main contribution of this paper is proposing an architecture for plenoptic photography -i.e., an architecture for sampling the full plenoptic function- by evaluating computational design parameters.

2. Compressive light field spectral imaging

The proposed architecture for compressive light field spectral imaging with coded apertures is composed by a main lens, a lenslet array, a coded aperture, a dispersive element (a prism), and a focal plane array (FPA) detector, as illustrated in figure 1.(a). Formally, let $f_0(x, y, u, v, \lambda)$ be the plenoptic function describing the light flow of a given scene. The image that is formed after the lenslet array, is given by [1]: $f_1(x, y, u, v, \lambda) = \iint f_0(x, y, u, v, \lambda) A(u, v) \cos^4 \theta du dv$, where $A(u, v)$ is the aperture function (e.g. one within the opening of the camera and zero outside it) and θ is the angle subtended by the microlens. Starting from the plenoptic function sampled by the lenslet array, each angular position of f_1 can be spatially modulated by an angular-dependent coded aperture $T(x, y, u, v)$. This coded aperture is applied to the spatio-spectral angular density $f_1(x, y, u, v, \lambda)$, resulting in the coded field $f_2(x, y, u, v, \lambda) = T(x, y, u, v) f_1(x, y, u, v, \lambda)$. This coded aperture remains fixed and every angle of the scene is modulated by a different pattern in the coded aperture. The coded intensity is then spectrally dispersed by a dispersive element before it impinges on the detector as $f_3(x, y, \lambda, u, v) = f_2(x, y, u, v, \lambda) * h(x, y, u, v, \lambda)$, where $*$ denotes convolution, and $h(x, y, u, v, \lambda)$ is the optical impulse response of the dispersive system that, ideally, corresponds to a wavelength-dependent spatial shift. The compressive measurements across the detector for the (x, y) spatial coordinates are obtained by integrating the field $f_3(x, y, u, v, \lambda)$ over the spectral range sensitivity of the camera, Λ , and the $\Delta x \times \Delta y$ pixel area of the detector as [5, 6]: $Y(x, y) = \int_{\Lambda} \int_{\Delta x} \int_{\Delta y} f_3(\hat{x}, \hat{y}, u, v, \lambda) d\lambda d\hat{x} d\hat{y}$. To describe the discrete mathematical model, the voxel $F_{i,j,k,m,n}$ represents a measure of the intensity concentrated in a specific angular and spatio-spectral region of the source irradiance, where i and j index the spatial coordinates ($i = 0, \dots, N - 1, j = 0, \dots, M - 1$), k determines the k^{th} spectral plane ($k = 0, \dots, L - 1$), and m and n index the angular position ($m = 0, \dots, U - 1, n = 0, \dots, V - 1$). This discretization yields

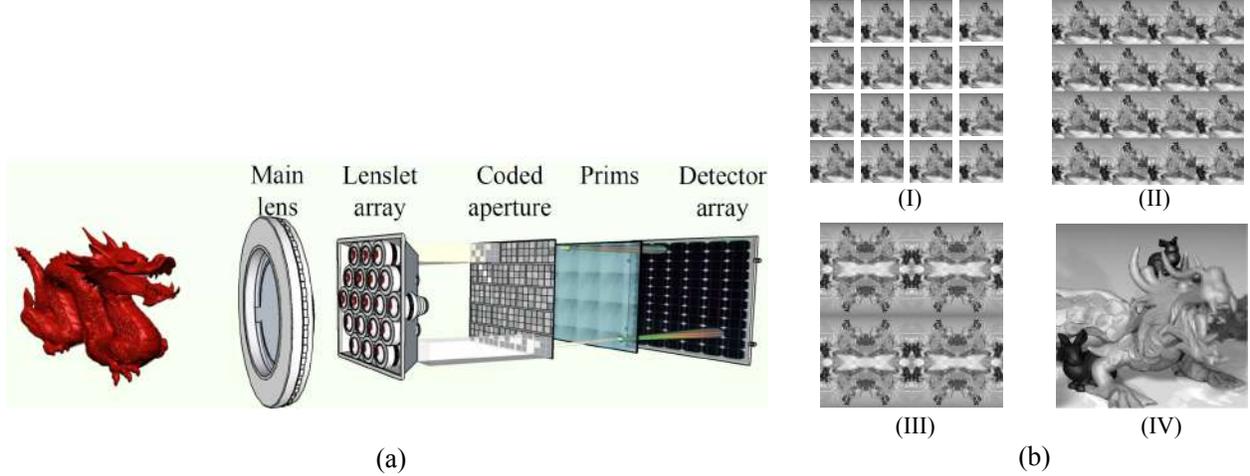


Fig. 1: (a) The proposed architecture exploits a microlens array behind the main lens of the camera combined with coded aperture patterns and a dispersive element. (b) Schematic representation illustrating the effects of rearranging the spectral light-field image: (I) unarranged, (II) mosaicking, (III) rotating, and (IV) macro.

a 5-dimensional discrete representation of the scene $\mathbf{F} \in \mathbb{R}^{N \times M \times L \times U \times V}$, where $N \times M$ is the number of spatial samples, L is the number of spectral bands and, $U \times V$ is the number of angular samples. Let $T_{i,j,m,n}$ be the discretized coded aperture, the energy captured on the detector that comes from the (m,n) -th angle, can be written as:

$$(Y_{i,j})_{m,n}^q = \sum_k F_{i,(j-k),k,m,n} T_{i,(j-k),m,n}^q + w_{(m,n,i,j)} \quad (1)$$

where the dispersion effect of the prism is represented by the shifting in the j -axis and, w is the noise in the detector. The measurement set acquired from a single angular position, $\mathbf{Y}_{m,n} \in \mathbb{R}^{M \times N \times L}$, can be represented in vector form as $\mathbf{y}_{m,n} \in \mathbb{R}^{MNL}$. Similarly, the spatio-spectral source \mathbf{F} can be expressed in vector form as $\mathbf{f} \in \mathbb{R}^{NMLUV}$, and the relation between the (m,n) -th source angular position and its correspondent measurement set is given by: $\mathbf{y}_{m,n} = \mathbf{H}_{m,n} \mathbf{f}_{m,n} + \mathbf{w}_{m,n}$, where $\mathbf{f}_{m,n}$ is the vector representation of the spatio-spectral source at the (m,n) -th angular position, and $\mathbf{H}_{m,n}$ is the single-shot sensing matrix that accounts for the effects of the aperture pattern $T_{i,j,m,n}$. Furthermore, measurements acquired from different angular positions can also be arranged in a single vector, $\mathbf{y} = [\mathbf{y}_{0,0}^T, \mathbf{y}_{1,0}^T, \dots, \mathbf{y}_{m-1,0}^T, \mathbf{y}_{0,1}^T, \mathbf{y}_{1,1}^T, \dots, \mathbf{y}_{m-1,n-1}^T]^T$. As a result, equation (1) can be expressed in matrix form as $\mathbf{y} = \mathbf{H}\mathbf{f} + \mathbf{w}$, where $\mathbf{H} \in \mathbb{R}^{UVM(N+L-1) \times MNLUV}$ is the single-shot sensing matrix for the complete angular scene. Specifically, the matrix \mathbf{H} is given by $\mathbf{H} = \text{diag}(\mathbf{H}_{0,0}, \mathbf{H}_{1,0}, \dots, \mathbf{H}_{m-1,0}, \mathbf{H}_{1,0}, \mathbf{H}_{1,1}, \dots, \mathbf{H}_{m-1,n-1})$. Typically, in compressive spectral imaging a single snapshot architecture allows the reconstruction of the underlying spectral data cube. However, following the work in [6] multiple snapshots using different coded aperture patterns yield a less ill-posed inverse problem, and better quality reconstructions. Consequently, equation (2) can be rewritten for dealing with multiple shots as: $\mathbf{y}_{m,n}^q = \mathbf{H}_{m,n}^q \mathbf{f} + \mathbf{w}_{m,n}^q$, where $\mathbf{y}_{m,n}^q$ corresponds to the q -th shot at angle (m,n) , for $q = 0, \dots, Q-1$, being Q the total number of shots. Each shot uses a different coded aperture $T_{i,j,m,n}^q$. All the measurement shots captured for a single angle can be arranged as $\mathbf{y}_{u,v} = [(\mathbf{y}_{u,v}^0})^T, (\mathbf{y}_{u,v}^1})^T, \dots, (\mathbf{y}_{u,v}^{Q-1})^T]^T$, with $\mathbf{y} = [\mathbf{y}_{0,0}^T, \mathbf{y}_{1,0}^T, \dots, \mathbf{y}_{m-1,0}^T, \mathbf{y}_{0,1}^T, \mathbf{y}_{1,1}^T, \dots, \mathbf{y}_{m-1,n-1}^T]^T$. Similarly, the multi-shot output can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{f} + \mathbf{w}, \quad (2)$$

where $\mathbf{H}^q = \text{diag}(\mathbf{H}_{0,0}, \mathbf{H}_{1,0}, \dots, \mathbf{H}_{1,0}, \mathbf{H}_{1,1}, \dots, \mathbf{H}_{m-1,n-1})$, and $\mathbf{H} = [(\mathbf{H}^0})^T, \dots, (\mathbf{H}^{Q-1})^T]^T$. Taking into account that the radiance in the same spatio-spectral position is expected to be correlated for different angles, this can be exploited by re-arranging the original data structure. Based on this concept, the 5D light field spectral representation is reduced to 3D. However, appropriate data structures for spectral light-field images have not been yet considered in the literature. In this work, three data structures have been proposed. Each data structure is obtained by rearranging the original (unarranged) structure \mathbf{F} in three different ways in order to obtain a 3D arrangement $\tilde{\mathbf{F}}$ as follows:

- Mosaicking structure: The spatio-spectral images are concatenated along the angular dimensions: $\tilde{\mathbf{F}}_{i,j,k} = \mathbf{F}_{a,b,k, [i/M], [j/N]}$, where $a = i - [i/M]M$ and $b = j - [j/N]N$ for $i = 0, \dots, (M-1)U$, and $j = 0, \dots, (N-1)V$ $\text{mod}(\cdot)$ denotes the module operation and $[\cdot]$ denotes round-off to the closest lower integer value.
- Rotating structure: Similarly as in the mosaicking arrangement, the spatio-spectral cubes are concatenated along the angular dimensions, but they are alternatively rotated and flipped in order to favor the continuity of spatial

SNR [dB]	Data Structure	Number of shots [Q]			
		1	2	3	4
10	Macro	24.45	26.05	27.17	28.79
	Unarranged	20.33	23.99	24.89	25.83
	Rotating	19.62	22.86	23.57	24.23
	Mosaicking	17.36	20.48	21.59	22.65
15	Macro	24.84	28.01	30.64	32.54
	Unarranged	22.13	24.83	26.76	28.34
	Rotating	21.29	23.65	25.49	27.68
	Mosaicking	19.65	21.35	23.49	25.67
20	Macro	26.04	29.22	32.47	33.44
	Unarranged	22.42	25.08	31.47	29.71
	Rotating	21.18	23.89	30.28	28.19
	Mosaicking	19.36	21.48	28.69	26.87

Table 1. Mean reconstruction PSNR in dB with spectral bands of $L=8$, numbers of shots $Q=\{1,2,3,4\}$, and $SNR=\{10,15,20\}$.

patterns: $\tilde{\mathbf{F}}_{i,j,k} = \mathbf{F}_{(M-1)c+(a(-1)^c),(N-1)d+(b(-1)^d),k,[i/M],[j/N]}$, where $c = \text{mod}(\lfloor i/m \rfloor, 2)$ and $d = \text{mod}(\lfloor j/n \rfloor, 2)$.

- Macro structure: The pixels in the same spatial and spectral position are concatenated along corresponding angular dimensions: $\tilde{\mathbf{F}}_{i,j,k} = \mathbf{F}_{i-\lfloor i/U \rfloor U, j-\lfloor j/V \rfloor V, k, \lfloor i/U \rfloor, \lfloor j/V \rfloor}$.

Examples of each data structure are illustrated in Fig. 1.(b), where (I) depicts the unarranged light field image \mathbf{F} , and (II)-(IV) the re-arranged light field image $\tilde{\mathbf{F}}$ by the proposed structures.

3. Reconstruction of light field spectral images

In this section the performance of the proposed architecture is assessed in terms of the number of shots and the noise in the acquired data. For this purpose, a measurement set is simulated using the multi-shot model described in (2). The simulated measurement set is then used as the input of a GPSR reconstruction algorithm in order to obtain an approximation of the original scene. The robustness of the reconstructions with respect to the effect of noise in the measurements is studied in Table 1, and illustrated in Figure 2. Each PSNR value in table 1 corresponds to the average of 5 experiments with random coded aperture patterns. As expected, Table 1 indicates that a higher number of shots and the macro structure results in higher quality reconstructions. Furthermore, a higher noise level results in lower reconstruction quality. Notwithstanding, reconstruction qualities of up to 28.79 [dB] can be achieved with the macro structure with a SNR as low as 10[dB].

4. Conclusion

The results obtained in this work suggest that simultaneously sampling each dimension of the plenoptic function is feasible. Future work should consider artifacts and distortion introduced by the optical elements, such as the main lens, the lenslet array and the diffractive element, as well as the effect of reduced light transmission induced by the coded aperture, in order to obtain a more accurate model of the image formation process in the proposed architecture prior to the development of the first prototype.

References

1. E. H. Adelson and J. Y. Wang, "Single lens stereo with a plenoptic camera," *IEEE transactions on pattern analysis and machine intelligence*, vol. 14, no. 2, pp. 99–106, 1992.
2. A. Veeraraghavan, R. Raskar, A. Agrawal, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.*, vol. 26, no. 3, p. 69, 2007.
3. S. D. Babacan, R. Ansorge, M. Luessi, P. R. Mataran, R. Molina, and A. K. Katsaggelos, "Compressive light field sensing," *IEEE Transactions on image processing*, vol. 21, no. 12, pp. 4746–4757, 2012.
4. K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, p. 46, 2013.
5. D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," *Applied Optics*, vol. 49, no. 36, pp. 6824–6833, 2010.
6. M. Gehm, R. John, D. Brady, R. Willett, and T. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Optics express*, vol. 15, no. 21, pp. 14013–14027, 2007.

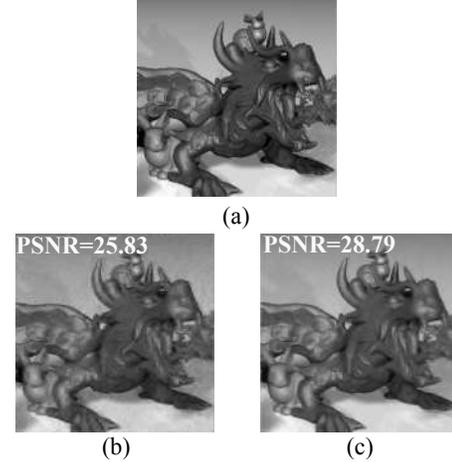


Fig. 2. Reconstruction of (a) using the (b) macro and (c) unarranged structure with fourth shots and $SNR=10$.