SLIC super-pixels for multi-resolution compressive spectral imaging reconstruction

Uso de super pixeles SLIC para la reconstrucción de imágenes espectrales adquiridas mediante muestreo compresivo

Hans Garcia¹, Claudia V. Correa² and Henry Arguello²*

1. Dept. of Electrical Engineering, 2. Dept. of Computer Science, Universidad Industrial de Santander, Colombia

(*) E-mail: hwarfu@uis.edu.co

ABSTRACT:
Spectral imaging (SI) is widely used in different applications involving material identification since it contains both spatial ($x, y$) and spectral information ($\lambda$). However, traditional SI scanning methods involve massive amounts of data, which increase the cost of storing and processing. Compressive sensing (CS) theory has been applied in SI, such that the underlying data cube can be recovered from a reduced number of measures. Reconstructions are obtained by $l_2 - l_1$ norm-based algorithms whose computational complexity grows in proportion to the number of unknowns. In this paper, a multi-resolution reconstruction model based on the simple linear iterative clustering (SLIC) is proposed to reduce the number of unknown values to recover. Simulation results show that the proposed method is up to 86% faster than the full-resolution reconstructions. Additionally, MR approach obtains more accurate reconstructions with improvements of up to 12dB of PSNR.

Key words: Multi-resolution, super-pixels, single pixel camera, compressive spectral imaging.

REFERENCES AND LINKS / REFERENCIAS Y ENLACES


1. Introduction

Several imaging applications require high resolution images for material classification tasks in areas such as astrophysics, environmental remote sensing, and biomedical image processing. In particular, spectral imagery (SI) is widely used in these applications since it contains both spatial \((x, y)\) and spectral information \((\lambda)\). The cost of acquiring SI using traditional scanning methods such as whisk-broom, push-broom or filtered cameras, directly depends on the specific application and the desired resolution, because they acquire the information of one or two dimensions and then scan the area of interest along the remaining dimension. Since they capture the information from all voxels in the scenes, the costs of storage and processing grow in proportion to the desired resolution [1]. On other hand, compressive sensing (CS) theory [2] has been recently applied to spectral imaging, which has been called compressive spectral imaging (CSI).

To date, several CSI optical architectures have been developed. Most of them employ a 2D intensity detector that captures the projections of the spectral scene, some examples include: the spatial-spectral encoded compressive spectral imager (SSCSI)[3], the coded aperture snapshot spectral imager (CASSI)[4], the dual-coded hyperspectral imager (DCSI)[5] and the prism-mask multispectral video imaging system (PMVIS)[6]. The implementation cost of these architectures, depends directly on the detector resolution required for each application[7]. A lower-cost alternative employs a single pixel camera (SPC), whose detector can be either a photodiode circuit or a point spectrometer (Whisk-broom) as in [8].

CSI establishes that it is possible to retrieve a spectral image \(\mathbf{F} \in \mathbb{R}^{M \times N \times L}\), where \(M\) and \(N\) are the spatial dimensions and \(L\) is the number of spectral bands, from a lower number of samples, under the assumption that the SI \(\mathbf{F}\) has a sparse representation in some basis. In particular, a vectorized form of a SI \(\mathbf{f} \in \mathbb{R}^{MNL}\) has a dispersion level \(S\), if it can be represented as a linear combination of \(S\) vectors on the basis \(\mathbf{Ψ}\), such that \(\mathbf{f} = \mathbf{Ψ} \mathbf{θ}\) with \(S \ll MNL\). Thus, instead of acquiring \(MNL\) samples, CS captures \(K \ll MNL\) projections of the scene, where \(K\) is not necessarily equal to the sparsity level \(S\). The sensing process can be represented in matrix form as \(\mathbf{g} = \mathbf{H} \mathbf{f}\), where \(\mathbf{H}\) is the transfer matrix of the optical system. Because the number of projections is considerably less than the number of variables to reconstruct, the problem is ill conditioned, and the reconstruction of the signal \(\mathbf{f}\) is obtained using optimization algorithms that take advantage of the sparse representation of \(\mathbf{f}\) in a transformation basis and solve the problem

\[
\hat{\mathbf{f}} = \mathbf{Ψ}^{-1} \left\{ \arg\min_{\mathbf{θ}} ||\mathbf{Ψ} \mathbf{θ} - \mathbf{g}||^2 + \tau ||\mathbf{θ}||_1 \right\},
\]

where \(\tau\) is a regularization parameter. Equation (1) is traditionally solved using iterative algorithms as in [9]–[12]. The cost of each iteration in these iterative reconstruction algorithms depends on the spatial dimensions of the object to reconstruct which is limited by the number of acquired measures.

In this work, a multi-resolution (MR) reconstruction model is proposed such that the computational complexity of the inverse problem is reduced by employing super-pixels, which in turn reduces the number of unknown values to recover in the inverse problem. Given that most scenes exhibit areas with similar spectral signatures, several pixels can be grouped into super-pixels without losing information or inducing considerable errors [12]–[15]. For instance, super-pixels have been previously used for gray-scale image segmentation [16]–[18]. The proposed model first obtains a fast-low-cost reconstruction of the scene, which is then used to construct the super-pixels by the single linear iterative clustering (SLIC) method. The obtained super-pixel map is used to generate the correspondent decimation matrix \(\mathbf{D}\) that modifies the minimization problem in Equation (1), such that, a MR sensing matrix is used instead of the full resolution matrix \(\mathbf{H}\). The new reconstruction optimization problem and the process to obtain \(\mathbf{D}\) are described in detail in the following sections. This paper is organized as follows: Section 2 presents the mathematical description of the single pixel camera employed in this work along with the proposed multi-resolution reconstruction scheme. Next, the use of SLIC super-pixels is described in Section 3. In order to evaluate the performance of the proposed methods, several simulation results are presented in Section 4, employing three different spectral scenes.
2. SPC Sensing and Multi-Resolution Reconstruction

The multi-resolution reconstruction approach proposed in this paper is based on the compressive spectral measurements acquired with a SPC, whose sensing process is illustrated in Fig. 1(a). This optical architecture is composed by an objective lens, a coded aperture $T \in \{0,1\}^{M \times N}$, a condenser lens and a whisk-broom spectrometer as a detector. Specifically, the objective lens focuses the light rays incoming from the scene to the coded aperture, which spatially modulates all the spectral bands. The coded scene is concentrated in the detector by the condenser lens[19]. Mathematically the acquired measurement for each spectral band $l$ is given by

$$g_l = \sum_i \sum_j F(i,j,l)T(i,j),$$

where $i = 1,2,\ldots,M$ and $j = 1,2,\ldots,N$. This model can be rewritten in matrix form as

$$g_l = h^T f_l,$$

for $l = 1,2,\ldots,L$, where $h$ is a vector form of $T$, given by $h = [T_{1,1}, T_{2,1}, \ldots, T_{M,N}]$ and $f_l$ is a vector form of the $l$-th spectral band of $F$. Furthermore, because only one measure is not sufficient to recover the SI, several snapshots $k = 1,\ldots,K$ are captured using different coded apertures $T^k$. The multi-shot sensing process for each band can be expressed as

$$g_l = H f_l,$$

where $H$ is a $K \times MN$ matrix, given by $H = [h_1^T, h_2^T, \ldots, h_K^T]^T$ where $h_k$ is the vector form of the different coded apertures $T^k$ used for each shot $k$, and $g_l = [(g_1^T)^T, (g_2^T)^T, \ldots, (g_L^T)^T]^T$. Furthermore, the sensing model for all the spectral bands and $K$ shots is given by

$$g = \tilde{H} f,$$

where $g = [(g_1^T)^T, (g_2^T)^T, \ldots, (g_L^T)^T]^T$, $\tilde{H}$ is the sensing matrix illustrated in Fig. 2 and can be obtained as a block diagonal matrix

$$\tilde{H} = \begin{bmatrix} H & 0 & \cdots & 0 \\ 0 & H & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H \end{bmatrix},$$

The compression ratio in this model is given by $\gamma = \frac{K}{MN}$, where the number of columns and rows of $\tilde{H}$ is equal to $MNL$ and $KL$, respectively. Notice that $\gamma \in [0,1]$. 

---

Fig. 1 (a) Single pixel camera (SPC) sensing phenomenon and (b) Proposed multi-resolution (MR) reconstruction process. Note that $\xi$ is calculated as the solution of the minimization problem in Equation (8).
Specifically, the sensing matrix $\mathbf{H}$ in this work is defined as $\mathbf{H} = \mathbf{WD} + \mathbf{Z}$ \cite{20}, where $\mathbf{W}$ is a $\gamma MN \times \gamma MN$ Hadamard matrix, $\mathbf{D}$ is a decimation matrix and $\mathbf{Z}$ is a $\gamma MN \times MN$ auxiliary binary matrix that improves the CS properties of $\mathbf{H}$. To obtain the image super-pixels, a fast low-resolution (LR) approximation of the spectral scene is calculated as the inverse Hadamard transform $\beta = \hat{\mathbf{W}}^T \mathbf{g}$, where $\hat{\mathbf{W}}$ is a block diagonal matrix with entries $\mathbf{W}$. Then, $\beta$ is interpolated to generate a full-resolution version of the scene $\hat{\mathbf{f}} = U\beta$. The interpolated scene is used to generate the MR downsampling matrix for each spectral band $\Delta$, which is related to the superpixel map. Notice that several MR downsampling matrices $\Delta$ can be used, this work proposes to build the matrix $\Delta$ based on image super-pixels using the simple linear iterative clustering (SLIC) algorithm \cite{4}. The process to generate $\Delta$ using the SLIC algorithm is explained in detail in section 3. Then, the down sampling matrix $\hat{\mathbf{D}}$ is built as a diagonal matrix whose entries correspond to the matrix $\Delta$, and is expressed as

$$
\hat{\mathbf{D}} = \begin{bmatrix}
[\Delta & 0 & \cdots & 0] \\
0 & [\Delta & \cdots & 0] \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & [\Delta] 
\end{bmatrix}.
$$

(7)

Including $\hat{\mathbf{D}}$ in the minimization problem from Eq.(1), and using the measures $\mathbf{g}$, the spectral signature for each super-pixel can be recovered as

$$
\xi = \Psi^{-1}\left\{\min_{\mathbf{\theta}} \|\hat{\mathbf{H}}\hat{\mathbf{D}}^T\Psi\mathbf{\theta} - \mathbf{g}\|_2^2 + \tau\|\mathbf{\theta}\|_1\right\},
$$

(8)

where $\tau$ is a regularization parameter and $\Psi$ is the sparse representation basis. Finally, using the spatial distribution of the super-pixels represented by $\hat{\mathbf{D}}$, the MR reconstructed S1 $\hat{\mathbf{f}}$ is given by

$$
\hat{\mathbf{f}} = \hat{\mathbf{D}}^T \xi.
$$

(9)

Algorithm 1 summarizes the described MR reconstruction process.

**Algorithm 1** Multi-resolution CSI reconstruction scheme

1: **Input:** Compressive measurements $\mathbf{g}$
2: **Output:** MR reconstructed S1 $\hat{\mathbf{f}}$
3: Reconstruct fast LR image $\beta = \hat{\mathbf{W}}^T \mathbf{g}$, where $\hat{\mathbf{W}}$ is a block diagonal matrix with entries $\mathbf{W}$.
4: Obtain interpolated fast full resolution image $\hat{\mathbf{f}} = U\beta$, where $U$ is a bilinear upsampling matrix.
5: Build $\Delta$ as explained in section 3.
6: Build MR decimation block diagonal matrix $\hat{\mathbf{D}}$ with entries $\Delta$.
7: Recover super-pixels solving $\xi = \Psi^{-1}\left\{\min_{\mathbf{\theta}} \|\hat{\mathbf{H}}\hat{\mathbf{D}}^T\Psi\mathbf{\theta} - \mathbf{g}\|_2^2 + \tau\|\mathbf{\theta}\|_1\right\}$
8: Recover full-resolution S1 $\hat{\mathbf{f}} = \hat{\mathbf{D}}^T \xi$

3. Design of Multi-Resolution Decimation Matrix (\Delta) Based on SLIC

The procedure to generate the MR matrix $\Delta$ is based on the analysis of the spectral signatures in the fast-full resolution reconstruction of the scene $\hat{\mathbf{f}}$. The main goal is to decompose the interpolated scene $\hat{\mathbf{F}}$ which is the 3D form of $\hat{\mathbf{f}}$ into subsets called super-pixels. There exist several methods for generating image super-pixels including graph-based, gradient accent, TurboPixels \cite{21}, quick super-pixels\cite{22}, and SLIC\cite{23}. This work employs the SLIC super-pixel method because of its low complexity $O(N)$, and high segmentation
accuracy[13]. It is important to note that SLIC super-pixels have been only used in RGB images, i.e. 3-band images. For this reason, we first calculate an equivalent 3-band version of the SI \( \mathbf{F} \) which may be obtained as

\[
\mathbf{f}_{(i,j,r)} = \frac{1}{\delta_L} \sum_{l=1}^{r-1} \delta_L \mathbf{f}_{(i,j,l)},
\]

where \( \delta_L = [L/3] \) and \( r = 1,2,3 \). Then, using the SLIC approach the super-pixels in this three-band version of the scene \( \mathbf{F} \) are determined. SLIC algorithm requires an input parameter \( \Gamma \), which indicates the number of equally-sized super-pixels of the resulting image. This parameter is used to determine the approximate size of the super-pixels as \( S = \sqrt{N/\Gamma} \). The main idea of the SLIC algorithm is to use the similarities of the pixels in a neighborhood of \( 2S \times 2S \) around of a center pixel, but the actual super-pixel size is only \( S \times S \). The center pixel is updated for each super-pixel until all pixels in the image have been assigned to a super-pixel. This algorithm is described with more detail in [23]. Fig. 3 illustrates the super-pixel grid for different number of super-pixels \( \Gamma \) and three data cubes. This figure suggests that the value of \( \Gamma \) determines the accuracy of the super-pixel map in preserving the spatial distribution of the scenes. More precisely, lower values of \( \Gamma \) result in a loss of detail, while larger values increase the amount of unknowns to recover.

![Image](https://www.sedoptica.es)

**Fig. 3** Visual comparison of the super-pixels map, using three different number of super-pixels \( \Gamma \) for the three test data cubes with original spatial resolution \( 120 \times 128 \) pixels.

The resulting super-pixel map is a matrix with integer entries \( P_{ij} \in \{1, \ldots, \Gamma\} \), for \( i = 1, \ldots, M, \; j = 1, \ldots, N \), where each value corresponds to a super-pixel index (label). The process to design the MR decimation matrix employs the super-pixel map. Specifically, the algorithm builds \( \Gamma \) binary matrices \( \Omega_s \) from \( P \), where \( s \) is the super-pixel index, such that for each matrix \( \Omega_s \) the pixels that belong to the \( s \)-th super-pixel take value 1 and the others zero, for each generated super-pixel. Mathematically, each \( \Omega_s \) can be obtained as

\[
(\Omega_s)_{ij} = \begin{cases} 
1, & P_{ij} = s \\
0, & \text{otherwise}.
\end{cases}
\]

Each binarized super-pixel map is vectorized as \( \text{vec}(\Omega_s) \) and added as a new row of the decimation matrix \( \Delta \), whose total number of rows is \( \Gamma \). An example of this process is illustrated in Fig. 4.

### 4. Simulations and Results

Several simulations were realized to test the performance of the proposed method when the number of bands \( L \) and the number of super-pixels \( \Gamma \) is changed, from 2 to 8 and 300 to 13000, respectively. Three different data cubes \( \mathbf{F} \) with \( 128 \times 128 \) pixels of spatial resolution and \( L = 8 \) spectral bands were used[24]. For each of these data cubes, the SPC compressive measures are obtained using the model in Eq. (5). In the traditional approach the SI is recovered solving the minimization problem in Eq. (1), using the gradient projection for sparse reconstruction algorithm (GPSR) [9]. The MR problem from Eq. (8) is also solved with the GPSR algorithm. In both cases the compression ratio was fixed to \( \gamma = 1/4 \) and the same number of iterations were used. The 3D representation basis \( \Psi = \Psi_3 @ \Psi_2D \) was selected as \( \Psi_3 \) the 1D discrete cosine transform and \( \Psi_2D \) a 2D Wavelet Symmlet 8 basis [25][26]. The comparisons are expressed in terms of peak signal to noise ratio (PSNR) and complexity/time of the reconstructions (seconds). All simulations were conducted and timed using an Intel Core i7-6700 @3.40GHz processor, and 32 GB RAM memory. Because dealing with the \( \mathbf{H} \) matrix for large images is a real issue, in practice our simulation code implements it as an operator. It is worth noting that the largest (complete) matrix \( \mathbf{H} \) from Eq. (6) accounts for all spectral
bands, and its dimensions depend on $M, N, L$ and the number of measurements. Because the same coding pattern is applied to all spectral bands, the implemented operator is related to $H$, such that it only depends on $M, N$ and the number of measurements. The implementation of this operator reduces the memory requirements.

![Diagram](image)

**Fig. 4** Example of the decimation matrix ($\Delta$) generation. White squares represent 1, and gray zones are 0. Three super-pixels are obtained from the $4 \times 4$ image.

### 4.a. Fast low-resolution reconstructions

The quality of the fast LR approximations $\beta$ is important to obtain a good super-pixel map to generate the MR matrix $\mathbf{D}$. Fig. 5 shows the RGB version of the LR reconstructions compared with a decimated version of the ground truth with a spatial resolution $64 \times 64$ for each spectral scene. In addition, the spectral responses of two points of each scene are shown in Fig. 5. Notice that the normalized spectral responses in all cases are close to the ground truth.

![Graphs](image)

**Fig. 5** Comparison between the low-resolution version of the ground-truth and the fast and LR approximation with spatial resolution $64 \times 64$ and $L = 7$, for three data cubes.
4.b. Multi-resolution reconstructions

The MR reconstructions for the three data cubes were obtained solving the minimization problem in Eq. (8), and the decimation matrices $\Delta$ presented in section 3. Fig. 6 shows the average reconstruction PSNR, and Fig. 7 shows the running time for the three data cubes using the traditional approach in Eq. (1) and the proposed MR. The results are presented as a function of the number of desired super-pixels for each number of spectral bands $L = \{3,...,8\}$. These figures show that the proposed MR reconstruction approach using SLIC improves the traditional full-resolution reconstructions, specifically, the best quality is obtained with $\Gamma$ close to 1200 in all simulated cases.

![Comparison of the MR and traditional approaches in terms of the reconstruction quality (PSNR) for the three data cubes varying the number of desired super-pixels $\Gamma$, and different amounts of spectral bands.](image)

Specifically, for the three data cubes (Feathers, woman and dog), the proposed MR approach provides the best reconstruction with up to 5dB, 9dB and 12dB improvements over the traditional approach, respectively. In addition, in relation to the computation time required for each approach, it can be noticed that the MR reconstructions using SLIC super-pixels are up to 86% faster than the full-resolution reconstructions. It is important to remark that these comparisons do not include the time required to generate the matrix $\Delta$. Furthermore, Fig. 8, Fig. 9 and Fig. 10 illustrate an RGB mapping of the attained reconstructions for the three data cubes, respectively, compared with the traditional full-resolution reconstruction from Eq. (1), and the ground truth. In addition, these figures present a comparison of the recovered spectral bands for each data cube, where the improvements of the MR approach are noticeable.
Fig. 7 Comparison of the MR and traditional approaches in terms of the running time for the three data cubes varying the number of desired super-pixels and number of spectral bands.

Fig. 8 Comparison between the ground truth with the different reconstruction approaches for data cube 1 with spatial dimensions $128 \times 128$ and $L = 7$, using $\Gamma = 1200$ super-pixels for the proposed approach. PSNR values are included for each image.
Fig. 9 Comparison between the ground truth with the different reconstruction approaches for data cube 2 with spatial dimensions $128 \times 128$ and $L = 7$, using $P = 1200$ super-pixels for the proposed approach. PSNR values are included for each image.

Fig. 10 Comparison between the ground truth with the different reconstruction approaches for data cube 3 with spatial dimensions $128 \times 128$ and $L = 7$, using $P = 1200$ super-pixels for the proposed approach. PSNR values are included for each image.

5. Conclusions

A methodology for multi-resolution reconstruction of spectral images from single pixel measurements has been proposed. Specifically, SLIC super-pixels are used to reduce the number of unknowns to recover. Simulation results show that if the amount of super-pixels is correctly selected in the SLIC algorithm, the proposed method provides more accurate reconstructions when compared to the traditional full resolution reconstruction approach. Furthermore, MR reconstructions using SLIC super-pixels result in an improvement of up to 11.8 dB of PSNR and up to 86% faster than the full-resolution reconstructions, with both approaches acquiring only the 25% of the data i.e. $\gamma = 0.25$. Specifically, in the case of the proposed approach, this amount of data is used to obtain the fast LR version of the scene, while the full-resolution approach uses it to recover the whole data cube.

Acknowledgements

Claudia V. Correa is funded by Universidad Industrial de Santander through the program to support postdoctoral researchers of the vice-presidency of research and extension.