

DISEÑO DE UN DICCIONARIO ESPACIO ESPECTRAL PARA LA  
REPRESENTACIÓN ESCASA DE IMÁGENES ESPECTRALES

KEVIN ADEMIR ARIAS ROJAS

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE CIENCIAS  
ESCUELA DE FÍSICA  
BUCARAMANGA  
2022

DISEÑO DE UN DICCIONARIO ESPACIO ESPECTRAL PARA LA  
REPRESENTACIÓN ESCASA DE IMÁGENES ESPECTRALES

KEVIN ADEMIR ARIAS ROJAS

Trabajo de grado para optar al título de  
Magíster en Matemática Aplicada

Director

PhD. Henry Arguello Fuentes

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE CIENCIAS

ESCUELA DE FÍSICA

BUCARAMANGA

2022

## DEDICATORIA

Los esfuerzos que dejan como resultado este libro y otros productos los dedico a mis padres, hermanos y compañera que en los momentos difíciles me dan la fuerza para continuar con mis objetivos y son mi luz guía. A mi director por su admirable dedicación y esfuerzo en la formación de una nueva generación de personas diferentes con impacto investigativo y personal. A mis amigos en el ámbito académico y personal que contribuyen con el desarrollo personal y apoyo moral a través de diferentes ámbitos de la vida.

Una dedicación especial a mis padres, porque a pesar de los momentos difíciles siempre trabajamos juntos para superarlos.

En forma general, gracias a todas las personas que me rodean porque son personas excelentes que han hecho mucho por mí.

***En este punto de la vida he aprendido que de lo único que nunca me arrepentiré es del esfuerzo realizado para lograr algo bueno.***

## CONTENIDO

	pág.
<b>INTRODUCCIÓN</b>	<b>13</b>
<b>1. MARCO TEÓRICO</b>	<b>19</b>
1.1. MODELO DE ADQUISICIÓN DE IMAGEN ESPECTRAL	19
1.2. TEORÍA DE REPRESENTACIÓN ESCASA	21
1.3. DICCIONARIOS ESPACIO-ESPECTRAL SEPARABLES	23
1.4. DICCIONARIO ESPACIO-ESPECTRAL CONJUNTO	32
<b>2. FORMULACIÓN DEL PROBLEMA</b>	<b>34</b>
<b>3. DICCIONARIO CONVOLUCIONAL BASADO EN PROCESAMIENTO LOCAL ESPACIAL</b>	<b>36</b>
3.1. Aprendizaje de diccionario convolucional	36
3.2. Enfoque convolucional localizado	38
<b>4. DICCIONARIO CONVOLUCIONAL ESPACIO-ESPECTRAL CONJUNTO</b>	<b>42</b>
4.1. Estimación de los elementos $s_i^{(j)}$	44
4.2. Aprendizaje de diccionario convolucional	46
4.3. Actualización de la variable dual $u_i$	47
<b>5. PROCESO DE RESTAURACIÓN DE IMÁGENES ESPECTRALES</b>	<b>49</b>
5.1. Eliminación de ruido	52
5.2. Borrosidad	53
5.3. Super-resolución	54
<b>6. EXPERIMENTOS NUMÉRICOS</b>	<b>55</b>

6.1. RESULTADOS NUMERICOS	56
6.1.1. Eliminación de ruido	57
6.1.2. Borrosidad	58
6.1.3. Super-resolucion	58
6.2. COMPLEJIDAD COMPUTACIONAL	59
6.3. RESTAURACION DE IMAGENES	61
<b>7. CONCLUSIONES</b>	<b>68</b>
<b>BIBLIOGRAFÍA</b>	<b>69</b>

## LISTA DE FIGURAS

	<b>pág.</b>
Figura 1. Ilustración de los tres escenarios de distorsión sobre la imagen (a) original. (b) En la eliminación de ruido se utilizó una relación señal a ruido (SNR) de $10[dB]$ . (c) Distorsión por borrosidad, donde se usó un filtro gaussiano de convolución con desviación estándar de 0.5. (d) Adicional al filtro gaussiano aplicado en (c) se realizó una reducción de la resolución con un factor de escalamiento $s = 4$ para el escenario de super-resolución.	21
Figura 2. Ilustración del procedimiento de aprendizaje para un diccionario espacio espectral separable y adaptativo.	26
Figura 3. Ilustración del proceso de adquisición de una escena espectral sujeta a diferentes factores de distorsión.	35
Figura 4. Estructura de un diccionario convolucional basado en el enfoque de procesamiento local.	39

- Figura 5. a) Entrenamiento del diccionario  $\Phi_L$ : La imagen espectral distorsionada  $\mathcal{Y}$  es dividida en subimágenes  $\mathcal{Y}_i \in \mathbb{R}^{N_a M_a L}$  para reducir la complejidad computacional. Posteriormente, el conjunto de subimágenes  $\mathcal{Y}_i$  es dividido en dos conjuntos de datos conformados por 70 % de las muestras para entrenamiento y 30 % de las muestras para evaluación. Así, los datos de entrenamiento son usados para estimar el diccionario espectral  $\Phi_L$  bajo el algoritmo propuesto basado en localidad y los datos de validación son empleados para soportar la calidad de su estimación. b) Codificación de la representación escasa: Una vez la fase de entrenamiento a finalizado, el diccionario resultante  $\Phi_L$  es empleado en el algoritmo de codificación escasa para estimar los coeficientes  $\varepsilon_i^{(j)}$  asociados al 30 % de los datos de evaluación restantes, lo que es más conocido como la etapa de evaluación. En forma conjunta, los coeficientes  $\varepsilon_i^{(j)}$  asociados a los datos de entrenamiento (70 %) son también estimados para calcular la imagen espectral  $\hat{X}$  sin distorsión. 51
- Figura 6. Comparación visual de los métodos bicubico, bilineal, A2SDLHSI, HySure, NPTSR y la propuesta para Super resolución usando el conjunto de datos Chikusei. 61
- Figura 7. Comparación visual de los métodos bicubico, bilineal, A2SDLHSI, HySure, NPTSR y la propuesta para Super resolución usando el conjunto de datos Washington DC MALL. 62
- Figura 8. Comparación visual de los métodos LRMR, LRTV, SSTV, LSSTV, LRTDTV, LRTDGS y la propuesta para la eliminación de ruido usando el conjunto de datos Chikusei. 63
- Figura 9. Comparación visual de los métodos LRMR, LRTV, SSTV, LSSTV, LRTDTV, LRTDGS y la propuesta para la eliminación de ruido usando el conjunto de datos Washington DC MALL. 64

- Figura 10. Comparación visual de los métodos HL, FPD, SSTV, DB-WLRTR, y la propuesta para la reducción de la borrosidad usando el conjunto de datos Chikusei. 65
- Figura 11. Comparación visual de los métodos HL, FPD, SSTV, DB-WLRTR, y la propuesta para la reducción de la borrosidad usando el conjunto de datos Washington DC MALL. 66

## LISTA DE TABLAS

	<b>pág.</b>
Tabla 1. Rendimiento de la imagen ruidosa, los métodos LRMR , LRTV, SSTV, LSSTV, LRTDTV, LRTDGS y el enfoque propuesto sobre el conjunto de datos Chikusei: PSNR, SSIM, FSIM, ERGAS y SAM.	57
Tabla 2. Rendimiento de la imagen ruidosa, los métodos LRMR, LRTV, SSTV, LSSTV, LRTDTV, LRTDGS y el enfoque propuesto sobre el conjunto de datos Washington DC Mall: PSNR, SSIM, FSIM, ERGAS y SAM.	57
Tabla 3. Rendimiento de los métodos HL, FPD, SSTV, DB-WLRTR y el enfoque propuesto sobre el conjunto de datos Chikusei: PSNR, SSIM, ERGAS y SAM.	58
Tabla 4. Rendimiento de los métodos HL, FPD, SSTV, DB-WLRTR y el enfoque propuesto sobre el conjunto de datos Washington DC MALL: PSNR, SSIM, ERGAS y SAM.	58
Tabla 5. Rendimiento de los métodos Bicubico, Bilineal, A2SDLHSI, HySure, NPTSR y el enfoque propuesto sobre el conjunto de datos Chikusei: PSNR, ERGAS, SAM y UIQI.	59
Tabla 6. Rendimiento de los métodos Bicubico, Bilineal, A2SDLHSI, HySure, NPTSR y el enfoque propuesto sobre el conjunto de datos Washington DC MALL: PSNR, ERGAS, SAM y UIQI.	59
Tabla 7. Tiempo de ejecución en segundos (s) para los métodos Bicubico, Bilineal, A2SDLHSI, HySure, NPTSR y el enfoque propuesto sobre los conjuntos de datos Chikusei y Washington DC MALL.	59

## RESUMEN

**TÍTULO:** DISEÑO DE UN DICCIONARIO ESPACIO ESPECTRAL PARA LA REPRESENTACIÓN ESCASA DE IMÁGENES ESPECTRALES. \*

**AUTOR:** KEVIN ADEMIR ARIAS ROJAS. \*\*

**PALABRAS CLAVE:** IMAGEN ESPECTRAL, DISTORSIÓN DE IMAGEN, DICCIONARIO ESPACIO ESPECTRAL, ESCASEZ.

### DESCRIPCIÓN:

Los sistemas para la adquisición de imágenes hiperespectrales (HS del inglés *hyperspectral imaging*) son útiles en un rango diverso de aplicaciones que implican tareas de detección y clasificación. En este proceso de adquisición existen problemas de distorsión de imagen, los cuales han sido abordados de forma computacional asumiendo escasez como un conocimiento previo para regularizar el espacio de solución. En esta dirección, una base de representación muy conocida tal como el diccionario espacio-espectral separable es aprendido con características espaciales y espectrales de imágenes naturales. Tradicionalmente, cada elemento del diccionario o átomo es calculado como un producto entre átomos de diccionarios espacial y espectral independientes. Específicamente, el diccionario final contiene átomos espacio-espectrales donde las características espaciales son aprendidas solo de muestras espaciales y de igual manera, las características espectrales solo de muestras espectrales. Sin embargo, el aprendizaje de un diccionario separable ignora la correlación entre la información espacial y espectral de imágenes naturales. Este trabajo propone un diccionario espacio-espectral conjunto donde cada uno de sus átomos son aprendidos a partir de información espacial y espectral conjunta. Es decir, cada muestra en la etapa de aprendizaje del diccionario está compuesta de información espacial y espectral de forma simultánea. Los resultados experimentales con datos reales muestran que el diccionario propuesto supera los diccionarios del estado del arte para obtener una representación más escasa aumentando la calidad de reconstrucción de la imagen.

---

\* Trabajo de investigación

\*\* Facultad de Ciencias. Escuela de Física. Director: PhD. Henry Arguello Fuentes.

**Objetivo general:** Desarrollar un algoritmo para el aprendizaje de un diccionario espacio espectral de alta escases enfocado a problemas de restauración de imagen hiperspectral.

**Objetivos específicos:**

- Modelar matemáticamente el sistema óptico asociado a diferentes problemas de restauración de imagen hiperspectral.
- Establecer el mejor escenario de representación sobre el cual el modelo de escases pueda explotarse de forma eficaz.
- Implementar un algoritmo para el aprendizaje del diccionario espacio espectral a partir de los datos observados.
- Implementar un algoritmo de restauración de imagen basado en escases a partir del diccionario aprendido.

## ABSTRACT

**TITLE:** DESIGN OF A SPATIAL-SPECTRAL DICTIONARY FOR SPARSE REPRESENTATION OF SPECTRAL IMAGES. \*

**AUTHOR:** KEVIN ADEMIR ARIAS ROJAS \*\*

**KEYWORDS:** SPECTRAL IMAGE, IMAGE DISTORTION, SPATIAL-SPECTRAL DICTIONARY, SPARSITY.

### **DESCRIPTION:**

Hyperspectral (HS) imaging systems are useful in a diverse range of applications that involve detection and classification tasks. Distortion problems in the acquisition process have been computationally addressed assuming sparsity as prior knowledge to regularize the solution space. In this way, well-known representation basis such as the separable spatial-spectral dictionary are learned with spatial and spectral features of natural images. Traditionally, each dictionary element or atom is calculated as a product between atoms of independent spatial and spectral dictionaries. Specifically, the separable trained dictionary contains spatial-spectral atoms where the spatial features are learned from only spatial samples and likewise, the spectral features from only spectral samples. However, separable dictionary learning ignores the correlation between the spatial and spectral information of natural images. This work propose a joint spatial-spectral dictionary where each of its atoms are learned from joint spatial and spectral information. That is, each sample in the dictionary learning stage is composed of spatial-spectral information simultaneously. Experimental results with real data show that the proposed dictionary outperforms the state-of-the-art dictionaries for sparser representation and therefore the reconstruction quality of the image.

---

\* Research work

\*\* Facultad de Ciencias. Escuela de Física. Director: PhD. Henry Arguello Fuentes.

## INTRODUCCIÓN

Los sistemas ópticos para la adquisición de imágenes hiperespectrales (HS del inglés *hyperspectral imaging*) adquieren una escena en cientos de bandas espectrales contiguas, cada una tomada en una longitud de onda específica <sup>1</sup>. Las imágenes hiperespectrales están caracterizadas por una alta resolución espectral, la cual permite la identificación exacta de los diferentes materiales contenidos en la escena de interés <sup>2</sup>. El análisis de la información espectral en imágenes HS ha permitido el desarrollo de varias aplicaciones en los campos de sensado remoto <sup>3</sup>, imágenes médicas<sup>4</sup> y astronomía<sup>5</sup>. Sin embargo, debido a razones tecnológicas y del entorno, las imágenes HS y en general espectrales pueden sufrir degradaciones en el proceso de adquisición. Por ejemplo, factores en el entorno como la ausencia de luz ocasionan ruido sobre la imagen adquirida debido a la baja potencia recibida en ciertos píxeles del sensor. De igual forma, las condiciones dinámicas como el movimiento del sensor o del objeto sobre el que se adquiere la imagen pueden producir un efecto de borrosidad principalmente en las regiones de la imagen con alta frecuencia. Por otra parte, aspectos tecnológicos como el muestreo limitado de los

---

<sup>1</sup> Chein-I Chang. *Hyperspectral data exploitation: theory and applications*. John Wiley & Sons, 2007.

<sup>2</sup> José M Bioucas-Dias y col. «Hyperspectral remote sensing data analysis and future challenges». En: *IEEE Geoscience and remote sensing magazine* 1.2 (2013), págs. 6-36.

<sup>3</sup> Michael E Schaepman y col. «Earth system science related imaging spectroscopyAn assessment». En: *Remote Sensing of Environment* 113 (2009), S123-S137.

<sup>4</sup> Svetlana V Panasyuk y col. «Medical hyperspectral imaging to facilitate residual tumor identification during surgery». En: *Cancer biology & therapy* 6.3 (2007), págs. 439-446.

<sup>5</sup> NM Förster Schreiber y col. «The SINS survey: SINFONI integral field spectroscopy of z 2 star-forming galaxies». En: *The Astrophysical Journal* 706.2 (2009), pág. 1364.

dispositivos de adquisición y la ausencia de resolución espacial o espectral del sensor ocasionan una baja estimación de la imagen <sup>6</sup>. En general, estas degradaciones ocasionan pérdida de detalle y distorsión espectral considerable sobre la escena afectando su rendimiento para las tareas de detección y clasificación. Por lo tanto, es de gran interés la formulación de técnicas de restauración de imágenes HS previo a la realización de cualquier tarea de aplicación <sup>67</sup>. Tradicionalmente, asunciones de escasez sobre imágenes naturales han sido impuestas en diferentes procesos de restauración para regularizar el espacio de solución <sup>8, 9, 6, 10</sup>. Matemáticamente, la imagen hiperespectral o espectral  $\mathcal{X} \in \mathbb{R}^{N \times M \times L}$  con dimensiones espaciales  $N \times M$  y  $L$  bandas espectrales puede ser representada de forma concisa bajo su forma vectorial como  $\mathbf{x} \in \mathbb{R}^{NML} = \Psi\alpha$  donde  $\alpha$  es la representación escasa con a lo máximo  $K$  elementos diferentes de cero,  $K \ll NML$  y  $\Phi$  es el dominio de escasez <sup>11</sup>. Este dominio puede ser seleccionado para ser un diccionario analítico tal como las transformadas ortogonales Wavelet, Fourier o Coseno las cuales proporcionan un nivel de representación generalizado. La ventaja de estas bases es la posibilidad de realizar una inversión exacta entre los dominios de escasez e imagen y por

- 
- <sup>6</sup> Ying Fu y col. «Adaptive spatial-spectral dictionary learning for hyperspectral image restoration». En: *International Journal of Computer Vision* 122.2 (2017), págs. 228-245.
- <sup>7</sup> Hongyan Zhang y col. «Hyperspectral image restoration using low-rank matrix recovery». En: *IEEE transactions on geoscience and remote sensing* 52.8 (2013), págs. 4729-4743.
- <sup>8</sup> Zhiyuan Zha y col. «Image restoration using joint patch-group-based sparse representation». En: *IEEE Transactions on Image Processing* 29 (2020), págs. 7735-7750.
- <sup>9</sup> Yong Chen y col. «Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition». En: *IEEE transactions on cybernetics* (2019).
- <sup>10</sup> Yang Xu y col. «Nonlocal patch tensor sparse representation for hyperspectral image super-resolution». En: *IEEE Transactions on Image Processing* 28.6 (2019), págs. 3034-3047.
- <sup>11</sup> Emmanuel J Candès y Michael B Wakin. «An introduction to compressive sampling». En: *IEEE signal processing magazine* 25.2 (2008), págs. 21-30.

ello, se han empleado en las pasadas décadas para diversas áreas de investigación<sup>12,13,14</sup>. También han sido construidos diccionarios con funciones base entrenadas a partir de un conjunto de datos adicionales previamente recolectados<sup>15,16</sup> o en una forma adaptativa a partir de solo los datos bajo observación<sup>17,18</sup>. De forma general, el entrenamiento de diccionarios consiste en calcular las funciones base también conocidas como átomos a partir de datos disponibles. Esto les permite obtener una representación escasa más concisa comparado a diccionarios analíticos para cierto conjunto de datos que siguen una distribución específica, lo que también es conocido como un diccionario de representación específica<sup>15</sup>. En esta categoría de diccionarios cabe resaltar la representación convolucional de imagen ampliamente usada en una variedad de tareas de distorsión cuando la imagen observada se

- 
- <sup>12</sup> Jian Cheng y col. «Remote sensing image fusion via wavelet transform and sparse representation». En: *ISPRS journal of photogrammetry and remote sensing* 104 (2015), págs. 158-173.
- <sup>13</sup> BK Shreyamsha Kumar, MNS Swamy y M Omair Ahmad. «Visual tracking using structural local DCT sparse appearance model with occlusion detection». En: *Multimedia Tools and Applications* 78.6 (2019), págs. 7243-7266.
- <sup>14</sup> Ganggang Dong y col. «Target recognition in SAR images via sparse representation in the frequency domain». En: *Pattern Recognition* 96 (2019), pág. 106972.
- <sup>15</sup> Michal Aharon, Michael Elad y Alfred Bruckstein. «K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation». En: *IEEE Transactions on signal processing* 54.11 (2006), págs. 4311-4322.
- <sup>16</sup> Mostafa Sadeghi, Massoud Babaie-Zadeh y Christian Jutten. «Dictionary learning for sparse representation: A novel approach». En: *IEEE Signal Processing Letters* 20.12 (2013), págs. 1195-1198.
- <sup>17</sup> Kevin Arias, Edwin Vargas y Henry Arguello. «Hyperspectral and Multispectral Image Fusion based on a Non-locally Centralized Sparse Model and Adaptive Spatial-Spectral Dictionaries». En: *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE. 2019, págs. 1-5.
- <sup>18</sup> Xiaorui Song, Lingda Wu y Hongxing Hao. «Hyperspectral image denoising base on adaptive sparse representation». En: *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*. IEEE. 2018, págs. 735-739.

compone de tres canales de color RGB (del inglés, *red, green and blue*)<sup>19</sup>. Una imagen bajo la representación convolucional es representada como una sumatoria de convoluciones entre diferentes filtros  $\phi_i$  y sus respectivos mapas de características  $\alpha_i$ ,  $\mathbf{x} = \sum_i \phi_i * \alpha_i$ . Específicamente, surge un gran interés en<sup>20</sup> donde se propone una dirección para operar de forma local sin perder el sentido de globalidad que impone dicha representación, a este enfoque se le conoce como codificación convolucional basada en procesamiento local. Adicionalmente, los diccionarios aprendidos pueden ser clasificados según la señal a representar y la formulación de sus átomos. Por ejemplo, según la señal de entrada existen diccionarios para representar de forma escasa imágenes bidimensionales (2D), imágenes tridimensionales (3D) y regiones de píxeles contiguos sobre la imagen también conocidas como parches (2D o 3D). Según la formulación de los átomos, estos pueden ser construidos para representar características espaciales, espacio-espectrales de forma separable y espacio-espectrales de forma conjunta<sup>21</sup>. Actualmente, el estado del arte ha dedicado sus esfuerzos en la representación de imágenes espectrales (3D) y en aprovechar de forma eficiente sus propiedades internas (como la correlación espacio-espectral). Específicamente, los átomos del diccionario son aprendidos en forma separable para representar características espacio-espectrales de imágenes espectrales en problemas de restauración de imágenes<sup>6, 17</sup>. Bajo este enfoque un átomo puede ser visto como un elemento 3D, el cual está construido para representar de forma escasa parches 3D  $\mathcal{X}_i \in \mathbb{R}^{P \times P \times L}$  con información espacial y espectral de la imagen, donde  $P \times P$  representa las dimensiones espaciales del parche,  $L$  el

---

<sup>19</sup> Cristina Garcia-Cardona y Brendt Wohlberg. «Convolutional dictionary learning». En: *arXiv pre-print arXiv:1709.02893* (2017).

<sup>20</sup> Vardan Papyan y col. «Convolutional dictionary learning via local processing». En: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, págs. 5296-5304.

<sup>21</sup> Bogdan Dumitrescu y Paul Irofti. *Dictionary learning algorithms and applications*. Springer, 2018.

número de bandas espectrales y  $i$  el índice del parche asociado al número de parches extraídos de la imagen  $\mathbf{X}$ . Dicho diccionario espacio-espectral  $\Phi$  es calculado de forma separable. Primero, un subdiccionario espacial  $\mathbf{S} \in \mathbb{R}^{P^2 \times P^2}$  es aprendido a partir de solo características espaciales. Luego, de forma independiente el subdiccionario  $\mathbf{U} \in \mathbb{R}^{L \times L}$  es aprendido de solo características espectrales y finalmente el diccionario completo es calculado como el producto Kronecker entre estas bases  $\Phi^{-1} = \mathbf{S} \otimes \mathbf{U}$  <sup>6</sup>. Para mitigar esa limitación y aprovechar la correlación espacio-espectral en imágenes espectrales, un diccionario sobre-completo que establece un mayor número de átomos que la dimensionalidad de la señal a representar es construido tal que cada átomo es aprendido a partir de información espacial y espectral de forma conjunta, lo cual permite representaciones más escasas que los métodos de entrenamiento de diccionario separables <sup>22</sup>. Sin embargo, hasta la fecha, no se ha construido un diccionario espacio-espectral conjunto que a su vez aproveche las ventajas que incorpora el uso de técnicas avanzadas de aprendizaje de diccionario. Entre las ventajas más relevantes, se destacan la reducción de la dimensionalidad en el diccionario tradicional sobre-completo y la inclusión de propiedades intrínsecas no exploradas sobre imágenes espectrales.

Este trabajo propone un diccionario espacio-espectral conjunto entrenado bajo el modelo de codificación escasa convolucional basado en procesamiento local. En el procesamiento local se incluye una nueva formulación que extiende la relación local-global ya conocida en el modelo de representación convolucional para entrenar los filtros de convolución mientras se opera localmente sobre parches de la imagen. El aprendizaje del diccionario espacio-espectral convolucional es realizado de forma adaptativa y para mantener dicho proceso como un problema tratable las muestras de entrenamiento son seleccionadas como secciones 3D de la imagen espectral,

---

<sup>22</sup> Xing Lin y col. «Spatial-spectral encoded compressive hyperspectral imaging». En: *ACM Transactions on Graphics (TOG)* 33.6 (2014), págs. 1-11.

donde cada sección se asume como independiente. De esta manera, los filtros del diccionario son entrenados para satisfacer todos los parches 3D de la imagen espectral bajo las ventajas de eficiencia del enfoque de procesamiento local.

Los resultados experimentales muestran que el diccionario propuesto logra obtener representaciones más escasas comparado con otros métodos del estado del arte.

Con base en este resultado, se plantea la restauración de las imágenes espectrales en tres escenarios de degradación: inclusión de ruido, borrosidad y baja resolución.

Para cada problema de degradación se obtuvieron ganancias de hasta de  $1[dB]$ ,  $2[dB]$ ,  $2[dB]$  en términos de relación señal a pico de ruido (PSNR, del inglés *peak signal-to-noise ratio*), respectivamente.

## 1. MARCO TEÓRICO

En este capítulo se introducen los temas relevantes del estado del arte relacionados a este trabajo de investigación, tales como el modelo de adquisición de imágenes espectrales, la teoría general de representación escasa y los diccionarios para inducir escasez propuestos en la literatura.

La notación matemática a usar a lo largo de este capítulo se define a continuación: Letra minúscula para denotar un valor constante ( $a$ ), letra mayúscula para denotar las dimensiones de un arreglo ( $A$ ), letra minúscula en negrilla para denotar un vector ( $\mathbf{a}$ ), letra mayúscula en negrilla para denotar una matriz ( $\mathbf{A}$ ) y letra mayúscula caligráfica para denotar un tensor o arreglo 3D ( $\mathcal{A}$ ).

### 1.1. MODELO DE ADQUISICIÓN DE IMAGEN ESPECTRAL

Sea  $\mathcal{X} \in \mathbb{R}^{N \times M \times L}$  una imagen HS con  $N \times M$  píxeles espaciales y  $L$  bandas espectrales. Matemáticamente, la adquisición de la imagen espectral es modelada de forma lineal como

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (1)$$

donde  $\mathbf{y}$  son las medidas observadas,  $\mathbf{H}$  es una matriz de degradación,  $\mathbf{x} \in \mathbb{R}^{NML}$  es la SI en su forma vectorial y  $\mathbf{n}$  es el término de ruido aditivo inherente al sistema de adquisición. Las medidas obtenidas en la Ec. 1 son también conocidas como medidas distorsionadas siendo  $\mathbf{H}$  el tipo de distorsión introducido. En el caso ideal, donde no existe degradación,  $\mathbf{H}$  esta dada por la matriz identidad  $\mathbf{H} = \mathbf{I}$ , el ruido aditivo es nulo  $\mathbf{n} = 0$  y las medidas obtenidas corresponden a la SI  $\mathbf{y} = \mathbf{x}$ .

A continuación, se presentan los escenarios para diferentes distorsiones de las imágenes HS y las respectivas dimensiones para las matrices  $\mathbf{y}$ ,  $\mathbf{H}$  y  $\mathbf{n}$ .

- Eliminación de ruido:

La inclusión de ruido en una escena es un suceso intrínseco del sistema de adquisición y de otros factores del entorno. Para este caso, la matriz de degradación  $\mathbf{H}$  es la identidad  $\mathbf{H} \in \mathbb{R}^{NML \times NML} = \mathbf{I}$  y el ruido del sensor es modelado por el vector  $\mathbf{n} \in \mathbb{R}^{NML}$ , tal que

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \in \mathbb{R}^{NML} \quad (2)$$

- Borrosidad:

La inclusión de borrosidad ocasiona la pérdida de detalles en una escena. Matemáticamente, la borrosidad para un pixel de la imagen puede modelarse como una ponderación entre pixeles vecinos y es denotada por la matriz  $\mathbf{H} \in \mathbb{R}^{NML \times MNL} = \mathbf{B}$ , donde  $\mathbf{B}$  es una matriz con estructura convolucional diseñada para aplicar un filtro específico sobre la imagen  $\mathbf{x}$ . Generalmente, para evaluar el modelo bajo la distorsión única de borrosidad se puede asumir ruido nulo  $\mathbf{n} \in \mathbb{R}^{MNL} = 0$ . De acuerdo con lo anterior, las medidas se calculan como

$$\mathbf{y} = \mathbf{B}\mathbf{x} \in \mathbb{R}^{MNL} \quad (3)$$

- Super-resolución:

El problema de super-resolución se presenta tradicionalmente cuando la imagen HS es adquirida usando un sistema que posee un sensor con baja resolución espacial. Por lo tanto, la matriz  $\mathbf{H} \in \mathbb{R}^{N_s M_s L \times NML} = \mathbf{PB}$  está compuesta por una matriz de borrosidad  $\mathbf{B}$  y una matriz  $\mathbf{P} \in \mathbb{R}^{N_s M_s L \times NML}$  que disminuye la resolución de la imagen, donde  $N_s = \frac{N}{s}$  y  $M_s = \frac{M}{s}$  serán las nuevas dimensiones espaciales de baja resolución reducidas por un factor de escala  $s \in \mathbb{N}$ .

Así, la imagen observada de baja resolución  $y$  está dada por

$$\mathbf{y} = \mathbf{P}\mathbf{B}\mathbf{x} \in \mathbb{R}^{N_s M_s L} \quad (4)$$

Note que en la Ec. 4 el vector de ruido es nulo, es decir,  $\mathbf{n} \in \mathbb{R}^{N_s M_s L} = 0$ , sin embargo, también es posible considerar el problema de super-resolución en condiciones donde se presente ruido en el sistema, es decir,  $\mathbf{n} \neq 0$ .

Los escenarios de distorsión de eliminación de ruido, borrosidad y super-resolución son presentados visualmente en la Figura 1

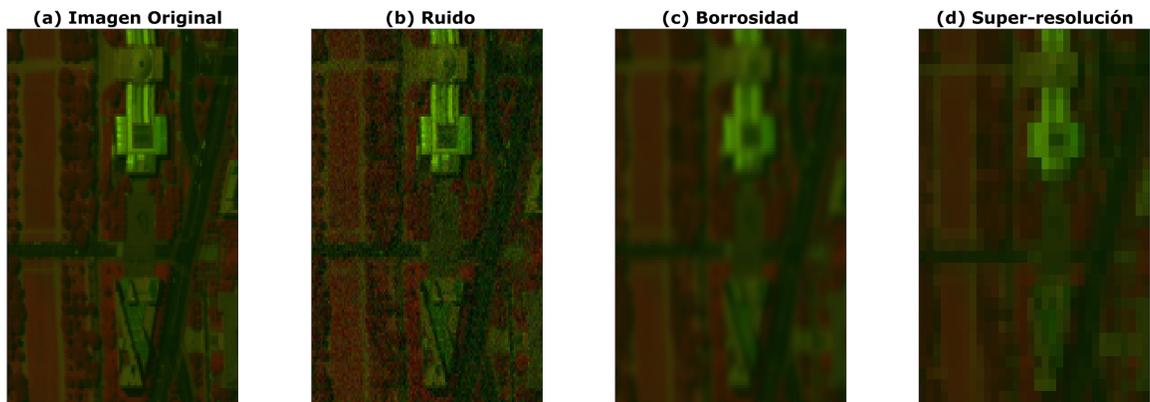


Figura 1. Ilustración de los tres escenarios de distorsión sobre la imagen (a) original. (b) En la eliminación de ruido se utilizó una relación señal a ruido (SNR) de  $10[dB]$ . (c) Distorsión por borrosidad, donde se usó un filtro gaussiano de convolución con desviación estándar de 0.5. (d) Adicional al filtro gaussiano aplicado en (c) se realizó una reducción de la resolución con un factor de escalamiento  $s = 4$  para el escenario de super-resolución.

## 1.2. TEORÍA DE REPRESENTACIÓN ESCASA

Estudios acerca del sistema visual humano sugieren que las imágenes naturales pueden ser representadas de forma escasa en algún dominio de transformación da-

do.<sup>23</sup> Los dominios para lograr escasez han sido clasificados como diccionarios (o bases) analíticos y aprendidos, los cuales contienen funciones base que representan primitivas estructurales de imagen convencional. Matemáticamente, una imagen natural en su representación vectorial  $\mathbf{x} \in \mathbb{R}^{NML}$  con  $N \times M$  píxeles de resolución espacial y  $L$  bandas espectrales puede ser representada de forma escasa por un número finito de átomos  $k \ll NML$

$$\mathbf{x} = \sum_{d=1}^k \Phi_d \alpha_d = \Phi \alpha \quad (5)$$

donde  $\Phi_d \in \mathbb{R}^{NML}$  es un átomo del diccionario con  $d \in \{1, 2, \dots, NML\}$  y  $\alpha_d$  es el peso asociado a dicho átomo. El vector de representación (o de pesos)  $\alpha$  corresponde al vector nulo con solo  $k$  de sus elementos ( $\alpha_d$ ) diferentes de cero. Por lo tanto,  $\alpha$  es llamado a ser escaso y  $k \ll NML$ . Por otra parte, el diccionario  $\Phi \in \mathbb{R}^{NML \times Z}$  puede ser seleccionado a ser completo cuando  $Z = NML$  o sobre-completo cuando  $Z > NML$ . Entre los diccionarios analíticos completos existen ejemplos muy conocidos como la transformada de coseno discreta (DCT), Fourier (FFT), y Wavelet (DwT). Por otro lado, respecto a los diccionarios aprendidos, los diccionarios sobre-completos son los más usados en la literatura ya que presentan una mejor representación respecto a los sobre-completos. Esto, debido al incremento en el número total de átomos de representación. Note que, la dimensión de la imagen puede llevar al escenario no deseado de usar para representación un diccionario de alta dimensión. Por lo tanto, se han empleado diferentes estrategias para mitigar esta posible limitación, por ejemplo diccionarios analíticos para la dimensión espacial  $NM$  han sido empleados de forma computacional para representar la imagen banda a banda a lo largo del espectro  $L$  evitando representar la imagen completa. Para ello,

---

<sup>23</sup> Haitao Yin, Shutao Li y Leyuan Fang. «Simultaneous image fusion and super-resolution using sparse representation». En: *Information Fusion* 14.3 (2013), págs. 229-240.

definiremos la forma matricial de la imagen  $\mathbf{x}$  como una matriz  $\mathbf{X} \in \mathbb{R}^{L \times NM}$ , donde cada banda espectral de la imagen se concatenan a lo largo de las filas, entonces la representación espacial de  $\mathbf{X}$  se escribe como

$$\mathbf{A} = \mathbf{X}\Phi_b^T \quad (6)$$

donde  $\mathbf{X}^{(i)} \in \mathbb{R}^{NM}$  es la fila  $i$ -ésima de la matriz y a su vez la banda espectral  $i$ -ésima de la imagen,  $\mathbf{A} \in \mathbb{R}^{L \times NM}$  es la representación escasa banda a banda de la imagen sobre la base espacial  $\Phi_b \in \mathbb{R}^{NM \times NM}$ . Generalmente,  $\Phi_b$  es seleccionada a ser la DWT.

### 1.3. DICCIONARIOS ESPACIO-ESPECTRAL SEPARABLES

En una forma separable, se ha incluido también un diccionario analítico  $\Phi_s \in \mathbb{R}^{L \times L}$  para representar el espectro de la imagen, donde tradicionalmente  $\Phi_s$  es seleccionada a ser la DCT. La representación espacial y espectral separable se expresa como

$$\mathbf{A} = \Phi_s \mathbf{X} \Phi_b^T \quad (7)$$

De acuerdo a las propiedades del producto Kronecker, la ecuación 7 puede reescribirse en forma vectorial como  $\alpha = (\Phi_b \otimes \Phi_s) \mathbf{x}$ . Para encontrar una relación con la ecuación 5 definimos la descomposición escasa de la imagen bajo un diccionario separable como  $\mathbf{x} = (\Phi_b^{-1} \otimes \Phi_s^{-1}) \alpha$ , donde  $\Phi = \Phi_b^{-1} \otimes \Phi_s^{-1}$ . El hecho de usar diccionarios analíticos separables y en la mayoría de los casos ortogonales permite reducir la complejidad y velocidad computacional para realizar las proyecciones directa e inversa sobre estas bases. Sin embargo, en diccionarios aprendidos la etapa de entrenamiento para la dimensión que contiene la información espacial  $NM$  puede llevar a un problema de alta computación. Por lo tanto, el espacio de carac-

terísticas espaciales de una imagen ha sido reducido al nivel de parches de menor dimensión. Diferentes diseños de diccionario espacio espectral separable son construidos para promover escasez, generalmente en problemas de restauración de imagen tales como fusión de imagen (usando dos imágenes, una multiespectral y una hiperespectral), super-resolución de imagen espectral (usando dos imágenes, una pancromática y una espectral) y cualquier otro problema en el cual se provean dos fuentes de imagen diferentes. Tradicionalmente, las dos fuentes corresponden a,  $\mathbf{Y}_M \in \mathbb{R}^{N \times M \times L_m}$  una imagen con máxima resolución espacial y  $\mathbf{Y}_H \in \mathbb{R}^{M_s \times N_s \times L}$  una imagen con máxima resolución espectral (ya sea una imagen espectral o una imagen hiperespectral). Matemáticamente, la adquisición de estas imágenes se modela como

$$\begin{aligned}\mathbf{Y}_M &= \bar{\mathbf{R}}\mathbf{X} + \mathbf{N}_M \\ \mathbf{Y}_H &= \mathbf{X}\bar{\mathbf{B}}\bar{\mathbf{P}} + \mathbf{N}_H\end{aligned}\tag{8}$$

donde  $\bar{\mathbf{B}}$  y  $\bar{\mathbf{P}}$  son las transformaciones de distorsión en modo matricial asociadas a  $\mathbf{B}$  y  $\mathbf{P}$  respectivamente,  $\mathbf{X} \in \mathbb{R}^{L \times NM}$  es la imagen  $\mathcal{X}$  en su forma matricial,  $\bar{\mathbf{R}} \in \mathbb{R}^{L_m \times L}$  es el operador de degradación espectral y las matrices  $\mathbf{N}_M$  y  $\mathbf{N}_H$  corresponden a los términos de ruido aditivo para las imágenes  $\mathbf{Y}_M$  y  $\mathbf{Y}_H$  respectivamente. Ahora, se supone la descomposición escasa separable de la  $\mathbf{X}$ , tal que

$$\mathbf{A} = \mathbf{U}\mathbf{X}\mathbf{S}^T\tag{9}$$

donde  $\mathbf{A}$  es la representación escasa en modo matricial,  $\mathbf{S} = \Phi_b$  es un diccionario con características base espaciales y  $\mathbf{U} = \Phi_s$  es un diccionario con características base espectrales. En una forma vectorial

$$\boldsymbol{\alpha} = (\mathbf{S} \otimes \mathbf{U}) \mathbf{x} = \Phi \mathbf{x}\tag{10}$$

De esta forma, las dos fuentes de información espacial y espectral pueden explotarse para aprovechar la complementariedad de su información <sup>17</sup>. Así la base espacial  $\mathbf{S}$  es construida a partir de una imagen con alta resolución espacial  $\mathbf{Y}_M$  y la base espectral  $\mathbf{U}$  a partir de una imagen con alta resolución espectral  $\mathbf{Y}_H$ . A continuación se mostrara las principales estrategias reportadas en el estado del arte para la construcción de las bases  $\mathbf{S}$  y  $\mathbf{U}$ .

- *Adaptive spatial-spectral dictionary learning for hyperspectral image restoration (2017)* <sup>6</sup>

En este trabajo, las bases espaciales y espectrales son construidas para representar de forma escasa parches 3D  $\mathcal{X}_i \in \mathbb{R}^{P \times P \times L}$  de la imagen mediante el análisis de componentes principales PCA (del inglés, *principal component analysis*). Específicamente, cada parche  $\mathbf{X}_i \in \mathbb{R}^{P^2 \times L}$  puede ser representado por un número pequeño de átomos de la siguiente forma <sup>6</sup>

$$\mathbf{X}_i = \sum_m \alpha_{i,m} \Phi_m = (\mathbf{S} \otimes \mathbf{U})^{-1} \alpha_i \quad (11)$$

donde  $\mathbf{S} \in \mathbb{R}^{P^2 \times P^2}$  y  $\mathbf{U} \in \mathbb{R}^{L \times L}$  son bases ortonormales que expanden las características espaciales y espectrales de la imagen, respectivamente. En el dominio espacial,  $h$  parches 2D son extraídos sobre la imagen con alta resolución espacial y reorganizados en una matrix  $P^2 \times h$  para aprender la base  $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{P^2}]$  usando PCA. En el dominio espectral, la imagen con alta resolución espectral es reorganizada en una matrix  $L \times NM$  y de igual forma PCA es empleado para obtener la base  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_L]$ . Para este trabajo, las bases de representación  $\mathbf{S}$  y  $\mathbf{U}$  son calculadas de forma adaptativa a partir de la estimación actual  $\bar{\mathcal{X}}$  de alta resolución espacial y espectral dentro de un algoritmo iterativo de optimización siguiendo el procedimiento que se muestra en la Figura 2. Note que, se conforman diferentes subdiccionarios  $\Phi_k$ , cada

uno asociado al grupo  $k$  de parches similares.

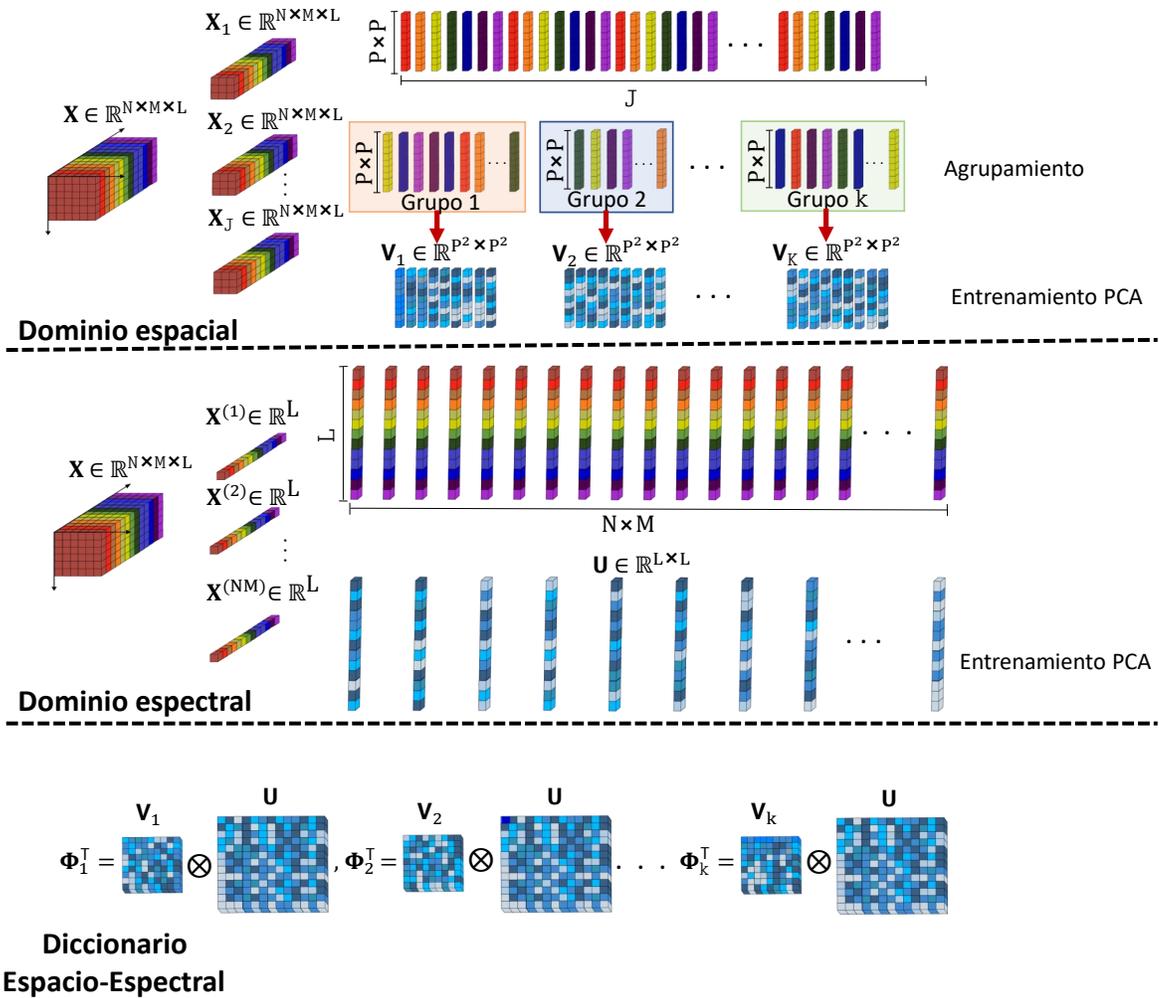


Figura 2. Ilustración del procedimiento de aprendizaje para un diccionario espacio espectral separable y adaptativo.

Otros trabajos en el estado del arte han construido un diccionario de escasez bajo el mismo enfoque de diseño separable con algunas variantes adicionales<sup>24, 17</sup>. Además, se ha propuesto el diseño de diccionarios espacio-espectrales separables

<sup>24</sup> Ying Fu y col. «Adaptive spatial-spectral dictionary learning for hyperspectral image denoising». En: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, págs. 343-351.

con enfoques diferentes al PCA tradicional.

- *Hyperspectral image super-resolution via non-local sparse tensor factorization* (2017) <sup>25</sup>.

Antes de introducir el método, se presentarán las siguientes definiciones tensoriales: Un tensor  $Z$ -dimensional es denotado por  $\mathcal{M} \in \mathbb{R}^{I_1, I_2, \dots, I_Z}$  con elementos  $m_{i_1, i_2, \dots, i_Z}$  donde  $1 \leq i_z \leq I_z$ . Los elementos desplegados en el modo- $z$  del tensor  $\mathcal{M}$  son los vectores de dimensión  $I_z$  extraídos a partir de  $\mathcal{M}$  variando el índice  $i_z$  mientras que los otros índices se mantienen fijos. La matriz desplegada en el modo- $z$  se denota como  $\mathbf{M}^{(z)} \in \mathbb{R}^{I_z \times I_1 \cdots I_{z-1} I_{z+1} \cdots I_Z}$  contiene todos los vectores modo- $z$  reorganizados como columnas de la matriz.

El producto de dos matrices puede generalizarse como la multiplicación entre un tensor y una matriz. El producto modo- $z$  entre un tensor  $\mathcal{M}$  y la matriz  $\mathbf{B} \in \mathbb{R}^{J_z \times I_z}$ , denotado por  $\mathcal{M} \times_z \mathbf{B}$  es un tensor  $Z$ -dimensional  $\mathcal{C} \in \mathbb{R}^{I_1 \times \dots \times J_z \times \dots \times I_Z}$  cuyos elementos son calculados como

$$c_{i_1 \dots i_{z-1} j_z i_{z+1} \dots i_Z} = \sum_{i_z} m_{i_1 \dots i_{z-1} i_z i_{z+1} \dots i_Z} b_{j_z i_z} \quad (12)$$

El producto en el modo  $z$ ,  $\mathcal{M} \times_z \mathbf{B}$  también puede ser calculado como  $\mathbf{C}^{(z)} = \mathbf{B}\mathbf{M}^{(z)}$ . En una serie de multiplicaciones con diferentes modos el orden no es relevante  $\mathcal{M} \times_m \mathbf{A} \times_z \mathbf{B} = \mathcal{M} \times_z \mathbf{B} \times_m \mathbf{A}$ . Si los modos de las multiplicaciones son los mismos  $\mathcal{M} \times_z \mathbf{A} \times_z \mathbf{B} = \mathcal{M} \times_z (\mathbf{B}\mathbf{A})$ .

En este enfoque, la imagen hiperespectral de alta resolución es representada

---

<sup>25</sup> Renwei Dian, Leyuan Fang y Shutao Li. «Hyperspectral image super-resolution via non-local sparse tensor factorization». En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, págs. 5344-5353.

como un producto tensorial entre un tensor de coeficientes y diccionarios

$$\mathcal{X} = \mathcal{A} \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times_3 \mathbf{U} \quad (13)$$

donde  $\mathbf{V}_1$ ,  $\mathbf{V}_2$  y  $\mathbf{U}$  denotan los diccionarios en los modos filas, columnas y banda, respectivamente. De igual forma las imágenes (HS) hiperespectral y (MS) multiespectral asociadas a la imagen de alta resolución  $\mathcal{X}$  son representadas en forma tensorial como

$$\begin{aligned} \mathcal{Y}_H &= \mathcal{X} \times_1 \bar{\mathbf{P}}_1 \times_2 \bar{\mathbf{P}}_2, \\ \mathcal{Y}_M &= \mathcal{X} \times_3 \bar{\mathbf{R}} \end{aligned} \quad (14)$$

donde  $\bar{\mathbf{P}}_1$  y  $\bar{\mathbf{P}}_2$  son las matrices de reducción de solución a los largo de los modos ancho y alto, respectivamente y  $\bar{\mathbf{R}}$  es la matriz de degradación espectral. Con el objetivo de reducir la complejidad del problema de aprendizaje de diccionario tensorial, el tensor imagen es dividido en parches tensoriales 3D clasificados en grupos de similitud  $\mathcal{X}_{k,j}$

$$\mathcal{X}_{k,j} = \mathcal{A}_{k,j} \times_1 \mathbf{V}_{1,k} \times_2 \mathbf{V}_{2,k} \times_3 \mathbf{U}_k \quad (15)$$

donde  $k$  denota el  $k$ -esimo grupo y  $j$  el  $j$ -esimo parche tensorial de dicho grupo. Así, las matrices  $\mathbf{V}_{1,k}$ ,  $\mathbf{V}_{2,k}$  y  $\mathbf{U}_k$  denotan diccionarios a nivel de parche asociados al grupo de similitud  $k$ . De acuerdo con la Ec. 14, los parches  $\mathcal{Y}_{M,(k,j)}$  del  $k$ -esimo grupo de la imagen MS pueden ser formulados como

$$\mathcal{Y}_{M,(k,j)} = \mathcal{X}_{k,j} \times_3 \bar{\mathbf{R}} \quad (16)$$

Luego, sustituyendo  $\mathcal{X}_{k,j}$  en la Ec. 16

$$\mathcal{Y}_{M,(k,j)} = \mathcal{A}_{k,j} \times_1 \mathbf{V}_{1,k} \times_2 \mathbf{V}_{2,k} \times_3 \mathbf{U}_k^* \quad (17)$$

donde  $\mathbf{U}_k^* = \bar{\mathbf{R}}\mathbf{U}_k$  es el diccionario espectral transformado. Desplegando los tensores  $\mathcal{Y}_{M,(k,j)}$  y  $\mathcal{C}_{k,j} = \mathcal{A}_{k,j} \times_2 \mathbf{V}_{2,k} \times_3 \mathbf{U}_k^*$  en el modo ancho, la Ec. 17 puede ser reescrita como

$$\mathcal{Y}_{M,(k,j)}^{(1)} = \mathbf{V}_{1,k} \times \mathcal{C}_{1,k}^{(1)} \quad (18)$$

Note que, cada columna en  $\mathbf{T}_{V_{1k}} = [\mathcal{Y}_{M,(k,0)}^{(1)}, \mathcal{Y}_{M,(k,1)}^{(1)}, \dots]$  de acuerdo con Ec. 18 puede ser representada como una combinación lineal de las columnas en la matriz  $\mathbf{V}_{1,k}$ . Por lo que, la estimación de la matriz  $\mathbf{V}_{1,k}$  puede ser interpretada como un problema de aprendizaje de diccionario con restricción de escasez dado por

$$\begin{aligned} \min_{\mathbf{V}_{1,k}, \mathbf{B}_{V_{1k}}} & \|\mathbf{T}_{V_{1k}} - \mathbf{V}_{1,k} \times \mathbf{B}_{V_{1k}}\|_F^2 \\ \text{s.t.} & \|\mathbf{B}_{V_{1k}}(:, i)\|_0 \leq k_{1V} \end{aligned} \quad (19)$$

donde  $\mathbf{B}_{V_{1k}}$  es una variable auxiliar para denotar la representación escasa asociada al diccionario  $\mathbf{V}_{1,k}$  y  $k_{1V}$  el número máximo de elementos diferentes de cero. Siguiendo un procedimiento similar, los diccionarios para el modo columnas y espectral pueden ser estimados a partir de los siguientes problemas de

optimización

$$\begin{aligned} \underset{\mathbf{V}_{2,k}, \mathbf{B}_{V_{2k}}}{\operatorname{argmin}} \quad & \|\mathbf{T}_{V_{2k}} - \mathbf{V}_{2,k} \times \mathbf{B}_{V_{2k}}\|_F^2 \\ \text{s.a.} \quad & \|\mathbf{B}_{V_{2k}}(:, i)\|_0 \leq k_{2V} \end{aligned} \quad (20)$$

$$\begin{aligned} \underset{\mathbf{U}_k, \mathbf{B}_{U_k}}{\operatorname{argmin}} \quad & \|\mathbf{T}_{U_k} - \mathbf{U}_k \times \mathbf{B}_{U_k}\|_F^2 \\ \text{s.a.} \quad & \|\mathbf{B}_{U_k}(:, i)\|_0 \leq k_U \end{aligned} \quad (21)$$

donde  $\mathbf{T}_{V_{2k}} = [\mathcal{Y}_{M,(k,0)}^{(2)}, \mathcal{Y}_{M,(k,1)}^{(2)}, \dots]$ ,  $\mathbf{T}_{U_k} = [\mathcal{Y}_{H,(k,0)}^{(3)}, \mathcal{Y}_{H,(k,1)}^{(3)}, \dots]$  son las matrices de muestras espaciales en el modo alto y espectrales para el entrenamiento de los diccionarios en los modos alto y espectral, respectivamente; y  $\mathbf{B}_{V_{2k}}$  y  $\mathbf{B}_{U_k}$  son las matrices de representación escasa con  $k_{2V}$  y  $k_U$  elementos diferentes de cero asociadas a los diccionarios  $\mathbf{V}_{2,k}$  y  $\mathbf{U}_k$ , respectivamente.

Otros métodos del estado del arte también han incluido el diseño de un diccionario espacio-espectral bajo el enfoque tensorial para representar de forma concisa la imagen espectral en sus tres modos<sup>10, 26</sup>. Principalmente, la representación escasa bajo el esquema tensorial ha sido la base para la solución a los problemas de restauración de imagen espectral.

- *Convolutional dictionary learning for multi-channel signals* (2018)<sup>27</sup>.

---

<sup>26</sup> Shutao Li y col. «Fusing hyperspectral and multispectral images via coupled sparse tensor factorization». En: *IEEE Transactions on Image Processing* 27.8 (2018), págs. 4118-4130.

<sup>27</sup> Cristina Garcia-Cardona y Brendt Wohlberg. «Convolutional dictionary learning for multi-channel signals». En: *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE. 2018, págs. 335-342.

El aprendizaje de diccionario convolucional ha recibido gran interés para la solución de problemas inversos en procesamiento de señales e imágenes bi-dimensionales. Sin embargo, un esfuerzo por extender esta técnica a señales multi-canal tal como las imágenes hiperespectrales ha permitido la construcción de un diccionario espacio-espectral que puede ser escalado a señales con gran cantidad de bandas espectrales. El problema de aprendizaje de diccionario convolucional es usualmente formulado como

$$\operatorname{argmin}_{\{\phi_i\}, \{\alpha_{i,k}\}} \frac{1}{2} \sum_k \left\| \sum_i \phi_i * \alpha_{i,k} - \mathbf{t}_k \right\|_2^2 + \lambda \sum_{i,k} \|\alpha_{i,k}\|_1 \quad (22)$$

donde  $\{\phi_i\}$  es el conjunto de  $I$  átomos (o filtros) que conforman el diccionario convolucional,  $\mathbf{t}_k$  es la  $k$ -ésima imagen de entrenamiento y  $\alpha_{i,k}$  son los mapas de coeficientes escasos.

Con el objetivo de realizar la extensión espacio-espectral del diccionario convolucional, el estado del arte denota las imágenes hiperespectrales como un conjunto de  $L$  bandas espectrales  $\{\mathbf{t}_l\}$  o en forma matricial como  $\mathbf{T} = (\mathbf{t}_0, \mathbf{t}_1, \dots, \mathbf{t}_{L-1})$ . En este caso, el diccionario global espacio-espectral es construido a partir de dos elementos. El Primero es un subdiccionario espacial, el cual consiste en un conjunto de  $H_s$  filtros  $\{\mathbf{v}_i\}$ . Mientras que el segundo es un subdiccionario matricial  $\mathbf{U}$  con dimensión  $L \times H_c$  empleado para representar la información espectral de la imagen. Bajo este diccionario, la representación escasa consiste del conjunto de mapas de coeficientes  $\{\alpha_{i,l_s}\}$  donde  $i \in \{0, 1, \dots, H_s - 1\}$  denota el filtro relevante en  $\{\mathbf{v}_i\}$  y  $l_s \in \{0, 1, \dots, H_c - 1\}$  indica la columna relevante en el subdiccionario  $\mathbf{U}$ . Así, el aprendizaje del diccionario espacio-

espectral incluyendo un elemento convolucional esta dado por

$$\operatorname{argmin}_{\{\mathbf{A}_k\}, \mathbf{V}, \mathbf{U}} \frac{1}{2} \sum_k \|\mathbf{V} \mathbf{A}_k \mathbf{U}^T - \mathbf{T}_k\|_F^2 + \lambda \sum_k \|\mathbf{A}_k\|_1 \quad (23)$$

donde  $\mathbf{V}_i$  es definido tal que  $\mathbf{V}_i \boldsymbol{\alpha}_{i,l_s} = \mathbf{v}_i * \boldsymbol{\alpha}_{i,l_s}$ ,  $\mathbf{V} = (\mathbf{V}_0, \mathbf{V}_1, \dots, \mathbf{V}_{H_s})$  y

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\alpha}_{0,0} & \boldsymbol{\alpha}_{0,1} & \cdots \\ \boldsymbol{\alpha}_{1,0} & \boldsymbol{\alpha}_{1,1} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (24)$$

Note que, las matrices de coeficientes  $\mathbf{A}$  es definida para una única imagen  $\mathbf{T}$  para simplificar la notación ya que el problema de aprendizaje de diccionario es independiente para cada muestra. Sin embargo, el problema en la Ec. 23 es formulado para un conjunto de  $K$  imágenes  $\{\mathbf{T}_k\}$  y sus respectivos mapas de coeficientes  $\{\mathbf{A}_k\}$ .

Otro diseño de diccionario convolucional también fue propuesto en <sup>28</sup> para intentar aprovechar la correlación espacio-espectral de imágenes naturales.

#### 1.4. DICCIONARIO ESPACIO-ESPECTRAL CONJUNTO

A diferencia de un diccionario espacio-espectral separable en donde su construcción depende del aprendizaje independiente de dos bases, una espacial y otra espectral, un diccionario espacio-espectral conjunto es aprendido a partir de muestras que contiene simultáneamente información espacial y espectral. Un diccionario de

---

<sup>28</sup> C Barajas-Solano y col. «Tridimensional Convolutional Sparse Coding of Spectral Images». En: *Hyperspectral Imaging and Sounding of the Environment*. Optical Society of America. 2019, HTu3B-5.

aprendizaje conjunto fue propuesto en el estado del arte en el área de adquisición compresiva de imágenes hiperespectrales <sup>22</sup>. En este trabajo, un diccionario espacio-espectral es aprendido a partir de un conjunto de muestras de entrenamiento conformado a partir de parches de la imagen espacio-espectral. El conjunto de parches de tamaño  $P \times P \times L$  de entrenamiento es extraído de forma aleatoria a partir de un conjunto de imágenes hiperespectrales previamente recolectadas. Matemáticamente, el problema de aprendizaje del diccionario es formulado como el siguiente problema de optimización

$$\underset{\Phi, \alpha}{\operatorname{argmin}} \|\mathbf{T} - \Phi\alpha\|_2^2; \quad s.t. \quad \|\alpha_i\| \leq k, \quad \forall_i = 1, 2, \dots, \quad (25)$$

donde  $\Phi \in \mathbb{R}^{N^2P \times Q}$  es un diccionario espacio-espectral sobre-completo con  $Q$  átomos,  $\mathbf{T} \in \mathbb{R}^{P^2L \times O}$  es el conjunto de entrenamiento compuesto de  $O$  parches espacio-espectrales,  $\alpha = [\alpha_1, \dots, \alpha_O] \in \mathbb{R}^{Q \times O}$  es la matriz de vectores  $\alpha_i$  en sus columnas con máximo  $k$  elementos diferentes de cero como lo establece la restricción del problema de optimización.

Note que, un átomo de diccionario es aprendido a partir de información espacial y espectral simultáneamente ya que una muestra de entrenamiento consiste en un parche de la imagen 3D. Sin embargo, el aprendizaje del diccionario es no adaptativo y requiere recolectar un conjunto grande de datos de entrenamiento adicionales previo a la reconstrucción de la imagen (o a cualquier otro proceso de restauración de imagen).

## 2. FORMULACIÓN DEL PROBLEMA

La adquisición de imágenes espectrales es una técnica para adquirir una imagen tridimensional que contiene la irradiancia de una escena en varias bandas espectrales contiguas, cada una asociada a una longitud de onda específica <sup>17</sup>, como se ilustra en la Figura 3. Mientras que una cámara ordinaria adquiere luz en tres bandas de longitud de onda del espectro visible, rojo, verde y azul (RGB), la imagen espectral puede abarcar la adquisición de bandas en espectro visible y/o fuera de este rango espectral tal como infrarrojo, ultravioleta, rayos X o alguna combinación de los anteriores <sup>29</sup>, lo cual extiende sus capacidades en las áreas de clasificación y detección. Matemáticamente, la escena  $\mathcal{X} \in \mathbb{R}^{N \times M \times L}$  puede ser adquirida por una cámara espectral basada en escaneo por punto, escaneo por línea, escaneo por banda o adquisición instantánea <sup>30</sup>, lo cual hace referencia al sistema de adquisición. Un enfoque tradicional para medir la exactitud de las técnicas de restauración de imagen es suponer que la imagen adquirida por dicho sistema es la imagen objetivo <sup>6</sup>. Luego, esta imagen es sometida a un proceso computacional de degradación (o degradaciones) para obtener la imagen medida  $\mathcal{Y}$ . En una forma vectorial se tiene que

$$\mathbf{y} = \mathbf{H}\mathbf{x} \quad (26)$$

---

<sup>29</sup> Gary A Shaw y Hsiaohua K Burke. «Spectral imaging for remote sensing». En: *Lincoln laboratory journal* 14.1 (2003), págs. 3-28.

<sup>30</sup> Yu Winston Wang y col. «Multiplexed optical imaging of tumor-directed nanoparticles: a review of imaging systems and approaches». En: *Nanotheranostics* 1.4 (2017), pág. 369.

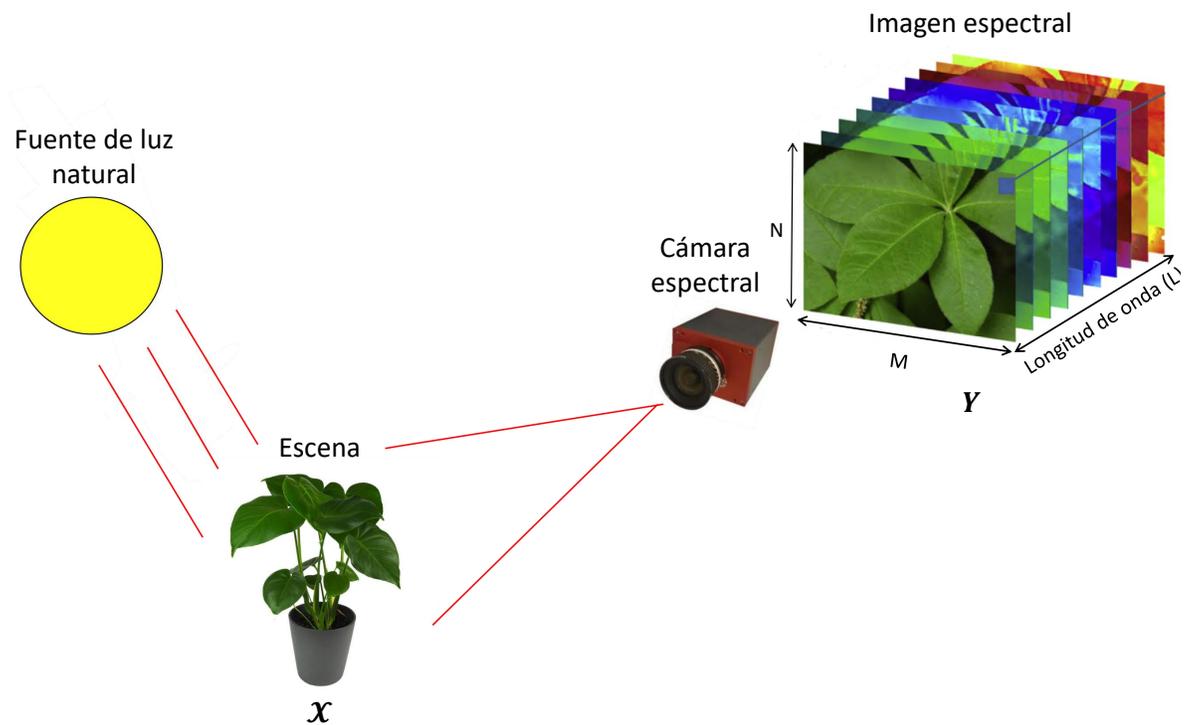


Figura 3. Ilustración del proceso de adquisición de una escena espectral sujeta a diferentes factores de distorsión.

donde  $\mathbf{H}$  es la matriz de degradación,  $\mathbf{x} \in \mathbb{R}^{NML}$  y  $\mathbf{y}$  corresponden a la representación vectorial de los tensores 3D  $\mathcal{X}$  y  $\mathcal{Y}$ , respectivamente. Note que, el sistema de adquisición en este trabajo es supuesto a ser la matriz identidad y la degradación  $\mathbf{H}$  de la imagen es el único proceso modelado para el problema de restauración abordado. Así, independientemente del sistema de adquisición, se desea obtener un diccionario de representación  $\Phi$  que induzca representaciones muy escasas  $\alpha$  de la imagen  $\mathbf{x} = \Phi\alpha$  para potenciar el escenario de regularización escasa y aumentar la exactitud de restauración de la imagen espectral. En las siguientes secciones se presentará el novedoso diccionario convolucional espacio-espectral conjunto para abarcar el objetivo mencionado.

### 3. DICCIONARIO CONVOLUCIONAL BASADO EN PROCESAMIENTO LOCAL ESPACIAL

Este capítulo presenta de forma general las bases teóricas para el planteamiento de un nuevo diccionario espacio-espectral conjunto que cumpla con los requerimientos de calidad de representación escasa y de recursos computacionales empleados.

#### 3.1. Aprendizaje de diccionario convolucional

La codificación escasa convolucional (CSC, del inglés *convolutional sparse coding*) sugiere un nuevo modelo de representación global para superar las limitaciones de representación escasa que impone el modelo tradicional basado en parches. Matemáticamente, la imagen  $\mathbf{x} \in \mathbb{R}^{NM(L=1)}$  puede ser descompuesta como

$$\mathbf{x} = \sum_{i=1}^m \phi_i * \alpha_i \quad (27)$$

donde  $\phi_i \in \mathbb{R}^Z$  son filtros locales de convolución que operan sobre sus respectivos mapas de características escasos  $\alpha_i \in \mathbb{R}^{NM(L=1)}$ ,  $(N, M)$  son las dimensiones espaciales de la imagen y  $L$  es la dimensión espectral o número de banda fijado como  $L = 1$ . Como aclaración, se removerá la variable  $L$  a lo largo de esta sección ya que no se ha formulado una representación convolucional de imágenes espectrales debido al costo computacional elevado que esto requiere, considerando que la representación opera sobre las imágenes de una forma global. De forma equivalente, el modelo en Ec. (27) puede reescribirse en su versión matricial como

$$\mathbf{x} = \Phi \alpha \quad (28)$$

donde  $\Phi \in \mathbb{R}^{NM \times (NM)m}$  es un diccionario convolucional en bandas con  $m$  filtros dife-

rentes y compuesto de versiones desplazadas del diccionario local  $\Phi_L = [\phi_1, \phi_1, \dots, \phi_m]$  y  $\alpha \in \mathbb{R}^{(NM)m}$  es la representación escasa conformada por una agregación de las representaciones  $\{\alpha_i\}_{i=1}^m$ . Bajo el modelo convolucional dado en la Ec. (28), la escena  $\mathbf{x}$  es una superposición de pocos átomos locales desplazados a diferentes posiciones en forma vertical, tal como se muestra en la Figura 4. Tradicionalmente, el aprendizaje de diccionario convolucional es direccionado bajo la solución del siguiente problema de optimización

$$\operatorname{argmin}_{\Phi, \Gamma} \frac{1}{2} \|\mathbf{X} - \Phi\Gamma\|_F^2 + \lambda \|\Gamma\|_1 \quad (29)$$

donde  $\mathbf{X} \in \mathbb{R}^{NM \times l} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l]$  es el conjunto de datos de entrenamiento compuesto de  $l$  imágenes y  $\Gamma \in \mathbb{R}^{(NM)m \times l} = [\alpha_1, \alpha_2, \dots, \alpha_l]$  son sus correspondientes vectores escasos. El problema de minimización en Ec. (29) es un problema compuesto de dos términos no convexos en las variables  $\Phi$  y  $\Gamma$  pero convexos cuando alguna de las dos variables se establece como una constante, ya sea convexo en  $\Phi$  con  $\Gamma$  constante o convexo en  $\Gamma$  con  $\Phi$  constante, donde un término se define como convexo de acuerdo a las garantías de convergencia hacia un mínimo, local o global, dependiendo del tipo de restricciones. Por lo tanto, debe emplearse un enfoque de solución alternando entre la minimización del diccionario  $\Phi$  y la representación escasa  $\Gamma$ , en el que cada uno de estos pasos recibe el nombre de actualización de diccionario y codificación escasa, respectivamente.

$$\begin{aligned} & \operatorname{argmin}_{\Gamma} \frac{1}{2} \|\mathbf{X} - \Phi\Gamma\|_F^2 + \lambda \|\Gamma\|_1, \\ & \operatorname{argmin}_{\Phi} \frac{1}{2} \|\mathbf{X} - \Phi\Gamma\|_F^2 + \iota_{C_{PN}}(\Phi) \end{aligned} \quad (30)$$

donde  $\iota_{C_{PN}}$  es una función indicador del conjunto de restricción  $C_{PN} = \{\forall i \mathbf{x}_i \in$

$\mathbb{R}^{NM} : \{\|\mathbf{x}_i\|_2 = 1\}$  dada por

$$\iota_S(X) = \begin{cases} 0, & \text{si } X \in S \\ \infty & \text{si } X \notin S \end{cases}$$

Algunas reformulaciones sobre los problemas en la Ec. (29) son llevadas a cabo para encontrar estructuras tradicionales de problemas de minimización que permitan el uso de algoritmos tradicionales en el campo de la representación escasa, tales como OMP (del inglés *convolutional orthogonal matching pursuit*), LARS (del inglés *generalized least angle regression*), K-SVD (del inglés *K-singular value decomposition*), MOD (del inglés *method of optimal directions*), aprendizaje de diccionarios en línea y *trainlets*. Adicionalmente, un mecanismo de acoplamiento debe ser empleado para solucionar el problema de optimización completo de la Ec. (30) y como se muestra en <sup>31</sup> existen diferentes estrategias para realizar este procedimiento. La elección de alguno de ellos debe ser estudiada de acuerdo al tipo de problema de aprendizaje de diccionario a solucionar.

### 3.2. Enfoque convolucional localizado

El aprendizaje de diccionario convolucional basado en procesamiento local es un enfoque novedoso propuesto en <sup>20</sup> para extender la relación local-global que pierden los diccionarios convolucionales tradicionales basando su solución en la formulación ADMM (del inglés *alternating direction method of multipliers*) en el dominio de Fourier. Matemáticamente, la señal global  $\mathbf{x}$  puede ser dividida en parches locales  $\mathbf{x}^{(j)} \in \mathbb{R}^{Z=P^2L}$  con una representación local que viene derivada de su configuración

---

<sup>31</sup> Cristina Garcia-Cardona y Brendt Wohlberg. «Convolutional dictionary learning: A comparative review and new algorithms». En: *IEEE Transactions on Computational Imaging* 4.3 (2018), págs. 366-381.

global, tal que  $\mathbf{x}^{(j)} = \Omega \boldsymbol{\gamma}^{(j)}$  donde  $\Omega \in \mathbb{R}^{Z \times (2Z-1)m}$  es una subsección del diccionario global  $\Phi$  y  $\boldsymbol{\gamma}^{(j)} \in \mathbb{R}^{(2Z-1)m}$  es el vector de representación escasa asociado a  $\mathbf{x}^{(j)}$  extraído del vector escaso global  $\boldsymbol{\alpha}$  como se muestra en la Figura 4.

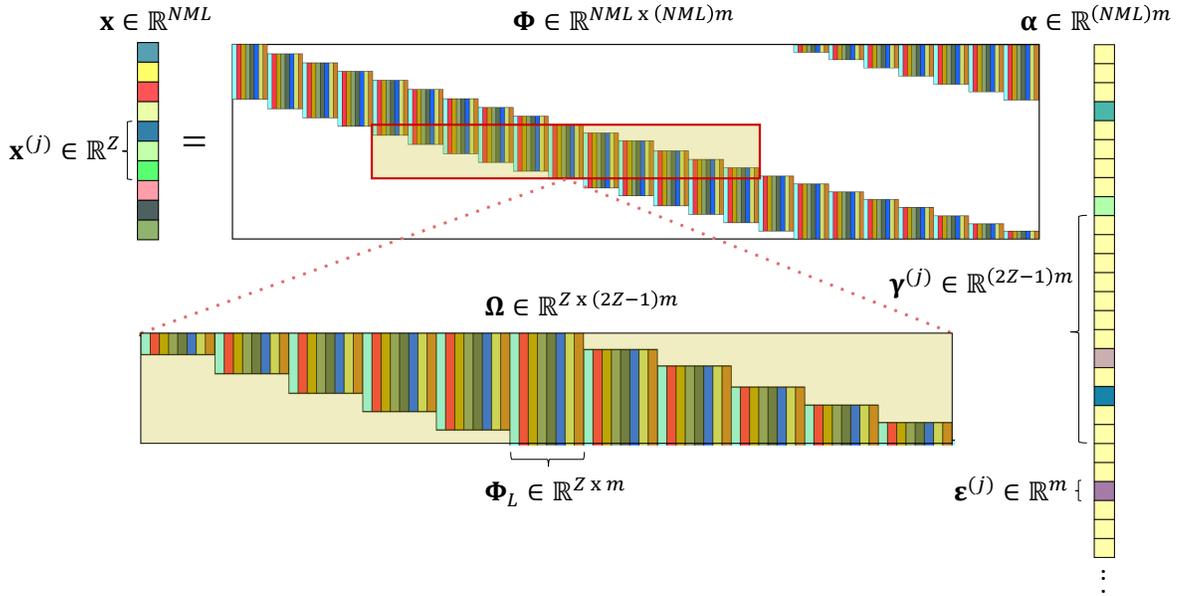


Figura 4. Estructura de un diccionario convolucional basado en el enfoque de procesamiento local.

El enfoque de representación localizado propuesto en el estado del arte solo ha sido abordado para imágenes acromáticas  $L = 1$ . Tenga en cuenta que suponer un modelo de representación localizado cuando la imagen es espectral o HS requiere un análisis por parches y una reformulación adicional, la cual será presentada en el siguiente capítulo para mitigar el impacto en los costos de almacenamiento. Por otra parte, la relación local-global en el problema de representación convolucional

ha sido analizada en los trabajos <sup>32</sup>, <sup>33</sup> en donde se establece que la solución del problema global es única y puede ser recuperada a partir de algoritmos *Greedy* cuando todos los elementos  $\gamma^{(j)}$  son escasos. En esta dirección, la reformulación basada en procesamiento local busca aprovechar esta relación local-global desde una perspectiva algorítmica, la cual permite solucionar de forma eficiente el problema global en función de solo operaciones locales. Como se muestra en la Figura 4, el vector escaso global  $\alpha$  puede ser dividido en un conjunto de vectores escasos  $\{\epsilon^{(j)} \in \mathbb{R}^m\}_{j=1}^{NML}$  y a su vez se establece la representación de la imagen global como

$$\mathbf{x} = \sum_{j=1}^{NML} \mathbf{D}_j^T \Phi_L \epsilon^{(j)} \quad (31)$$

donde  $\mathbf{D} \in \mathbb{R}^{Z \times NML}$  es el operador de extracción de parches de imagen tal que  $\mathbf{D}_j$  extrae el parche en la posición  $j$ ,  $\mathbf{x}^{(j)} = \mathbf{D}_j \mathbf{x}$  y  $\mathbf{D}_j^T \in \mathbb{R}^{NML \times Z}$  denota el operador de extracción transpuesto que fija el parche  $\Phi_L \epsilon^{(j)}$  en la posición  $i$  sobre un vector cero de dimensión  $NML$ . Ahora, incluyendo una nueva variable  $s_j = \Phi_L \epsilon^{(j)}$  se plantea el problema para entrenar el diccionario simplificado  $\Phi_L$  en lugar de  $\Omega$ . Por lo tanto, la imagen  $\mathbf{x}$  es ahora calculada por  $\mathbf{x} = \sum_{i=j}^{NML} \mathbf{D}_j^T s^{(j)}$  implicando que el algoritmo centre su análisis sobre los elementos  $s^{(j)}$  en lugar de los parches extraídos de la imagen  $\mathbf{x}^{(j)} = \mathbf{D}_j \mathbf{x}$  como se realiza tradicionalmente. Note que, los elementos  $s^{(j)}$  son entidades de estructura más simple y nos son parches extraídos de la imagen sino que un parche está constituido de elementos  $s_j$  superpuestos  $\mathbf{D}_j \mathbf{x} = \mathbf{D}_j \sum_{j=1}^{NML} \mathbf{D}_j^T s^{(j)}$ . Esta nueva formulación es motivada en <sup>20</sup> bajo el hecho que parches de imagen  $\mathbf{x}^{(j)}$  y sus respectivos elementos  $s^{(j)}$  son visualizados al

---

<sup>32</sup> Vardan Pappyan, Jeremias Sulam y Michael Elad. «Working locally thinking globally: Theoretical guarantees for convolutional sparse coding». En: *IEEE Transactions on Signal Processing* 65.21 (2017), págs. 5687-5701.

<sup>33</sup> Vardan Pappyan, Jeremias Sulam y Michael Elad. «Working locally thinking globally-part II: Stability and algorithms for convolutional sparse coding». En: *arXiv preprint arXiv:1607.02009* (2016).

mismo tiempo concluyendo que varios de estos elementos son más simples y por lo tanto contienen menos información. Como es bien conocido, este tipo de características son de gran interés para el área de representación escasa ya que estructuras más simples en  $s^{(j)}$  requieren un menor número de elementos de representación. La ventaja más relevante en el entrenamiento del diccionario convolucional basado en procesamiento local es encontrar la solución del problema global en función de solo cálculos locales bajo una estrategia localizada. Específicamente, la ejecución del algoritmo opera sobre parches de la imagen, pero soluciona fielmente para un contexto global. Entre otras ventajas importantes que incorpora la nueva formulación local se destaca la facilidad de escalabilidad a cualquier problema de restauración de imagen debido a su fácil implementación, rápida convergencia hacia la solución. Adicionalmente, este provee un mejor modelo de representación comparado con técnicas del estado del arte permitiendo representar los parches de imagen a partir de un número de valores diferentes de cero, una característica de interés cuando se analizan imágenes que poseen una estructura compleja y variante.

#### 4. DICCIONARIO CONVOLUCIONAL ESPACIO-ESPECTRAL CONJUNTO

En este capítulo se presenta el planteamiento del diccionario convolucional para la representación escasa de imágenes espectrales  $\mathbf{x} \in \mathbb{R}^{NML}$  con un número de bandas espectrales  $L \gg 3$ . El principal desafío del planteamiento del diccionario convolucional espacio-espectral conjunto recae en que el diccionario aproveche tanto la información espacial como la información espectral a lo largo de las  $L$  bandas, mientras que se cumple con requerimientos de calidad de representación y de recursos computacionales empleados. Para esto, en primer lugar, se opta por el enfoque de representación global que ofrece el diccionario convolucional evitando las limitaciones de representación que implica el uso de una estrategia local basada en parches. Segundo, se abarca el enfoque de división de imagen, tal que  $\mathbf{x}$  es dividida en  $N_s$  subimágenes  $\mathbf{x}_i \in \mathbb{R}^{N_d M_d L} = \mathbf{F}_i \mathbf{x}$ , donde  $\mathbf{F}_i \in \mathbb{R}^{N_d M_d L \times NML}$  es el operador de división de imagen que obtiene secciones no superpuestas de la imagen (las cuales serán supuestas a seguir el enfoque de procesamiento local mostrado en el capítulo anterior) para evitar el alto costo de almacenamiento y procesamiento que implicaría suponer la imagen espectral como una entidad global. En términos de implementación, el algoritmo soluciona para cada entidad  $\mathbf{s}_i^{(j)}$  asociada a la subimagen  $\mathbf{x}_i$  y al parche extraído en la posición  $j$  operando de forma local sin perder la relación con la solución global.

Por lo tanto, el problema de optimización general para el aprendizaje del diccionario convolucional espacio-espectral propuesto es formulado como

$$\begin{aligned}
 (\Phi_L, \mathbf{s}_i) = & \underset{\Phi_L, \{\boldsymbol{\varepsilon}_i^{(j)}\}_{j=1}^{N_d M_d L}, \{\mathbf{s}_i^{(j)}\}_{j=1}^{N_d M_d L}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{x}_i - \sum_{j=1}^{N_d M_d L} \mathbf{D}_j^T \mathbf{s}_i^{(j)}\|_2^2 + \lambda \sum_{j=1}^{N_d M_d L} \|\boldsymbol{\varepsilon}_i^{(j)}\|_1, \\
 \text{s.a. } \mathbf{s}_i^{(j)} = & \Phi_L \boldsymbol{\varepsilon}_i^{(j)}
 \end{aligned} \tag{32}$$

donde  $\mathbf{D}_j^T$  denota el operador de extracción que obtiene la subdivisión de la imagen en la posición  $j$ . La formulación en la Ec. (32) es un problema de optimización compuesto de dos términos, el primero corresponde a la fidelidad de los datos que busca la mejor aproximación entre el conjunto de entidades  $\{\mathbf{s}_i^{(j)}\}$  y los datos  $\mathbf{x}_i$ , y el segundo término que induce escasez sobre los vectores de representación de entidad  $\{\boldsymbol{\varepsilon}_i^{(j)}\}$ . Adicionalmente, el problema incorpora una restricción de igualdad  $\mathbf{s}_i^{(j)} = \Phi_L \boldsymbol{\varepsilon}_i^{(j)}$  para establecer la relación entre las entidades  $\{\mathbf{s}_i^{(j)}\}$  y sus respectivos vectores escasos  $\{\boldsymbol{\varepsilon}_i^{(j)}\}$ . Como ya fue mencionado, este tipo de problemas pueden ser solucionados desde una perspectiva alternante en donde el problema general se divide en subproblemas más manejables, y cada subproblema, asociado a una variable específica, es solucionado de forma independiente mientras los demás se establecen constantes. Por lo tanto, el problema en la Ec. (32) puede ser resuelto empleando el algoritmo ADMM (del inglés, *alternating direction method of multipliers*) encontrando la solución al Lagrangiano aumentado en su forma extendida que se expresa como

$$\begin{aligned}
(\Phi_L, \mathbf{s}_i, \mathbf{u}_i) = & \underset{\Phi_L, \{\boldsymbol{\varepsilon}_i^{(j)}\}_{j=1}^{N_d M_d L}, \{\mathbf{s}_i^{(j)}\}_{j=1}^{N_d M_d L}, \{\mathbf{u}_i^{(j)}\}_{j=1}^{N_d M_d L}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{x}_i - \sum_{j=1}^{N_d M_d L} \mathbf{D}_j^T \mathbf{s}_i^{(j)}\|_2^2 \\
& + \sum_{j=1}^{N_d M_d L} \left( \lambda \|\boldsymbol{\varepsilon}_i^{(j)}\|_1 + \frac{\rho}{2} \|\mathbf{s}_i^{(j)} - \Phi_L \boldsymbol{\varepsilon}_i^{(j)} + \mathbf{u}_i^{(j)}\|_2^2 \right), \tag{33}
\end{aligned}$$

La optimización de la función Lagrangiana es solucionada usando optimización alternante, lo cual se conforma de tres pasos: la estimación de los elementos  $\mathbf{s}_i^{(j)}$ , el aprendizaje tradicional de diccionario y la solución de la variable dual  $\mathbf{u}_i^{(j)} = \mathbf{u}_i^{(j)} + \mathbf{s}_i^{(j)} - \Phi_L \boldsymbol{\varepsilon}_i^{(j)}$ .

#### 4.1. Estimación de los elementos $\mathbf{s}_i^{(j)}$

$$\mathbf{s}_i = \underset{\{\mathbf{s}_i^{(j)}\}_{j=1}^{N_d M_d L}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{x}_i - \sum_{j=1}^{N_d M_d L} \mathbf{D}_j^T \mathbf{s}_i^{(j)}\|_2^2 + \frac{\rho}{2} \sum_{j=1}^{N_d M_d L} \|\mathbf{s}_i^{(j)} - \Phi_L \boldsymbol{\varepsilon}_i^{(j)} + \mathbf{u}_i^{(j)}\|_2^2 \quad (34)$$

Antes de entrar en la solución del problema reescribiremos la formulación en la Ec. (34) para expresar de forma explícita las sumatorias como

$$\begin{aligned} \mathbf{s} = \underset{\{\{\mathbf{s}_i^{(j)}\}_{j=1}^{N_d M_d L}\}_{i=1}^{N_s}}{\operatorname{argmin}} & \frac{1}{2} [\|\mathbf{x}_1 - (\mathbf{D}_1^T \mathbf{s}_1^{(1)} + \mathbf{D}_2^T \mathbf{s}_1^{(2)} + \dots + \mathbf{D}_{N_d M_d L}^T \mathbf{s}_1^{(N_d M_d L)})\|_2^2 \\ & + \|\mathbf{x}_2 - (\mathbf{D}_1^T \mathbf{s}_2^{(1)} + \mathbf{D}_2^T \mathbf{s}_2^{(2)} + \dots + \mathbf{D}_{N_d M_d L}^T \mathbf{s}_2^{(N_d M_d L)})\|_2^2 \\ & + \dots + \|\mathbf{x}_{N_s} - (\mathbf{D}_1^T \mathbf{s}_{N_s}^{(1)} + \mathbf{D}_2^T \mathbf{s}_{N_s}^{(2)} + \dots + \mathbf{D}_{N_d M_d L}^T \mathbf{s}_{N_s}^{(N_d M_d L)})\|_2^2] \\ & + \frac{\rho}{2} [\|\mathbf{s}_1^{(1)} - \Phi_L \boldsymbol{\varepsilon}_1^{(1)} + \mathbf{u}_1^{(1)}\|_2^2 + \|\mathbf{s}_1^{(2)} - \Phi_L \boldsymbol{\varepsilon}_1^{(2)} + \mathbf{u}_1^{(2)}\|_2^2 \\ & + \dots + \|\mathbf{s}_1^{(N_d M_d L)} - \Phi_L \boldsymbol{\varepsilon}_1^{(N_d M_d L)} + \mathbf{u}_1^{(N_d M_d L)}\|_2^2 + \dots + \\ & \|\mathbf{s}_{N_s}^{(N_d M_d L)} - \Phi_L \boldsymbol{\varepsilon}_{N_s}^{(N_d M_d L)} + \mathbf{u}_{N_s}^{(N_d M_d L)}\|_2^2] \end{aligned} \quad (35)$$

La solución al problema de optimización es llevada a cabo calculando su derivada con respecto a cada una de las variables  $\mathbf{s}_1^1, \mathbf{s}_1^2, \dots, \mathbf{s}_1^{N_d M_d L}, \dots, \mathbf{s}_{N_s}^{N_d M_d L}$  e igualando a cero para encontrar un mínimo sobre la función de optimización. Luego de realizar dicho procedimiento sobre el problema en la Ec. (35), se puede observar la

conformación del siguiente sistema de ecuaciones

$$\begin{aligned}
& -\mathbf{D}_1 \left( \mathbf{x}_1 - \sum_{j=1}^{N_d M_d L} \mathbf{D}_j^T \mathbf{s}_1^{(j)} \right) + \rho \left( \mathbf{s}_1^{(1)} - \Phi_L \boldsymbol{\varepsilon}_1^{(1)} + \mathbf{u}_1^{(1)} \right) = 0 \\
& -\mathbf{D}_2 \left( \mathbf{x}_1 - \sum_{j=1}^{N_d M_d L} \mathbf{D}_j^T \mathbf{s}_1^{(j)} \right) + \rho \left( \mathbf{s}_1^{(2)} - \Phi_L \boldsymbol{\varepsilon}_1^{(2)} + \mathbf{u}_2^{(2)} \right) = 0 \\
& \quad \vdots \\
& -\mathbf{D}_{N_d M_d L} \left( \mathbf{x}_1 - \sum_{i=1}^{N_d M_d L} \mathbf{D}_j^T \mathbf{s}_1^{(j)} \right) + \rho \left( \mathbf{s}_1^{(N_d M_d L)} - \Phi_L \boldsymbol{\varepsilon}_1^{(N_d M_d L)} + \mathbf{u}_1^{(N_d M_d L)} \right) = 0 \\
& \quad \vdots \\
& -\mathbf{D}_{N_d M_d L} \left( \mathbf{x}_{N_s} - \sum_{j=1}^{N_d M_d L} \mathbf{D}_j^T \mathbf{s}_{N_s}^{(j)} \right) + \rho \left( \mathbf{s}_{N_s}^{(N_d M_d L)} - \Phi_L \boldsymbol{\varepsilon}_{N_s}^{(N_d M_d L)} + \mathbf{u}_{N_s}^{(N_d M_d L)} \right) = 0,
\end{aligned} \tag{36}$$

El sistema de ecuaciones puede consolidarse realizando una concatenación de sus elementos. Así, definimos las siguientes matrices de concatenación

$$\begin{aligned}
\bar{\mathbf{D}} &= \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \vdots \\ \mathbf{D}_{N_d M_d L} \end{bmatrix}, \quad \bar{\mathbf{S}} = \begin{bmatrix} \mathbf{s}_1^{(1)} & \mathbf{s}_2^{(1)} & \cdots & \mathbf{s}_{N_s}^{(1)} \\ \mathbf{s}_1^{(2)} & \mathbf{s}_2^{(2)} & \cdots & \mathbf{s}_{N_s}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{s}_1^{(N_d M_d L)} & \mathbf{s}_2^{(N_d M_d L)} & \cdots & \mathbf{s}_{N_s}^{(N_d M_d L)} \end{bmatrix}, \quad \bar{\mathbf{Z}} = \\
& \begin{bmatrix} \Phi_L \boldsymbol{\varepsilon}_1^{(1)} - \mathbf{u}_1^{(1)} & \Phi_L \boldsymbol{\varepsilon}_2^{(1)} - \mathbf{u}_2^{(1)} & \cdots & \Phi_L \boldsymbol{\varepsilon}_{N_s}^{(1)} - \mathbf{u}_{N_s}^{(1)} \\ \Phi_L \boldsymbol{\varepsilon}_1^{(2)} - \mathbf{u}_1^{(2)} & \Phi_L \boldsymbol{\varepsilon}_2^{(2)} - \mathbf{u}_2^{(2)} & \cdots & \Phi_L \boldsymbol{\varepsilon}_{N_s}^{(2)} - \mathbf{u}_{N_s}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_L \boldsymbol{\varepsilon}_1^{(N_d M_d L)} - \mathbf{u}_1^{(N_d M_d L)} & \Phi_L \boldsymbol{\varepsilon}_2^{(N_d M_d L)} - \mathbf{u}_2^{(N_d M_d L)} & \cdots & \Phi_L \boldsymbol{\varepsilon}_{N_s}^{(N_d M_d L)} - \mathbf{u}_{N_s}^{(N_d M_d L)} \end{bmatrix}. \\
\bar{\mathbf{X}} &= \left[ \mathbf{x}_1^{(1)}, \mathbf{x}_2, \cdots, \mathbf{x}_{N_s} \right],
\end{aligned}$$

Remplazando en la Ec. (36) las nuevas variables  $\bar{\mathbf{R}}$ ,  $\bar{\mathbf{S}}$  y  $\bar{\mathbf{Z}}$  y solucionando para  $\bar{\mathbf{S}}$

obtenemos la solución al problema en la Ec. (34) dada por

$$\begin{aligned} -\bar{\mathbf{D}}(\bar{\mathbf{X}} - \bar{\mathbf{D}}^T \bar{\mathbf{S}}) + \rho(\bar{\mathbf{S}} - \bar{\mathbf{Z}}) &= 0 \\ \bar{\mathbf{S}} &= (\bar{\mathbf{D}}\bar{\mathbf{D}}^T + \rho\mathbf{I})^{-1}(\bar{\mathbf{D}}\bar{\mathbf{X}} + \rho\bar{\mathbf{Z}}), \end{aligned} \quad (37)$$

donde  $\mathbf{I} \in \mathbb{R}^{N_d M_d L \times N_d M_d L}$  es la matriz identidad. Con el fin de evitar el alto costo computacional en el cálculo inverso de  $(\bar{\mathbf{R}}\bar{\mathbf{R}}^T + \rho\mathbf{I})^{-1}$ , la solución en la Ec. (37) puede reescribirse bajo la identidad de Woodbury como

$$\begin{aligned} \bar{\mathbf{S}} &= \left( \frac{1}{\rho}\mathbf{I} - \frac{1}{\rho^2}\bar{\mathbf{D}} \left( \mathbf{I} + \frac{1}{\rho}\bar{\mathbf{D}}^T \bar{\mathbf{D}} \right)^{-1} \bar{\mathbf{D}}^T \right) (\bar{\mathbf{D}}\bar{\mathbf{X}} + \rho\bar{\mathbf{Z}}), \\ \bar{\mathbf{S}} &= (\mathbf{I} - \bar{\mathbf{D}}(\rho\mathbf{I} + n\mathbf{I})^{-1}\bar{\mathbf{D}}^T) \left( \frac{1}{\rho}\bar{\mathbf{D}}\bar{\mathbf{X}} + \bar{\mathbf{Z}} \right) \end{aligned} \quad (38)$$

La solución final en la Ec. (38) es simplificada bajo la siguiente identidad  $\bar{\mathbf{D}}^T \bar{\mathbf{D}} = \sum_{i=1}^{NML} \mathbf{R}_i^T \mathbf{R}_i = n\mathbf{I}$ . Note que, el uso de la formula Woodbury evita realizar un cálculo inverso costoso y en vez de ello el costo se reduce a calcular la transpuesta de una matriz.

## 4.2. Aprendizaje de diccionario convolucional

$$\Phi_L = \underset{\Phi_L, \{\{\epsilon_i^{(j)}\}_{j=1}^{N_d M_d L}\}_{i=1}^{N_s}}{\operatorname{argmin}} \sum_{i=1}^{N_s} \sum_{j=1}^{N_d M_d L} \left( \lambda \|\epsilon_i^{(j)}\|_1 + \frac{\rho}{2} \|\mathbf{s}_i^{(j)} - \Phi_L \epsilon_i^{(j)} + \mathbf{u}_i^{(j)}\|_2^2 \right) \quad (39)$$

El subproblema en la Ec. (39) bien conocido como aprendizaje de diccionario es abordado con un enfoque alternante en donde se actualizan los coeficientes  $\epsilon_i^{(j)}$  fijando el diccionario  $\Phi_L$  a ser constante y luego fijando los coeficientes  $\epsilon_i^{(j)}$  para

actualizar  $\Phi_L$ . específicamente, la minimización del problema de optimización en la Ec. (39) es separable con respecto a los coeficientes  $\{\{\epsilon^{(i)}\}_{j=1}^{N_d M_d L}\}_{i=1}^{N_s}$  y puede ser solucionado de forma independiente para cada  $\epsilon_i^{(j)}$ . Así, la estimación de los coeficientes bajo una restricción de escasez será llevada a cabo por el algoritmo estándar LARS (del inglés *Generalized least angle regression*)<sup>34</sup>. Luego, la minimización con respecto a  $\Phi_L$  puede ser realizada por cualquier algoritmo de aprendizaje de diccionario basado en parches tal como K-SVD, MOD, aprendizaje de diccionario en línea, entre otros. Solo unas pocas iteraciones de dicho algoritmo sin llegar a la convergencia son empleadas ya que en cada paso de alternancia los coeficientes cambian. Este algoritmo realiza de igual forma una actualización de los coeficientes  $\epsilon_i^{(j)}$  sujeto a una restricción de soporte que impone la salida del algoritmo de estimación de coeficientes.

### 4.3. Actualización de la variable dual $u_i$

Las variables duales  $u_i^{(j)}$  en el problema ADMM crecen hasta que han forzado el consenso entre las variables que se tratan por separado ( $s_i^j$  y  $\{\Phi_L, \epsilon_i^{(j)}\}$ ) para facilitar su cálculo. En el algoritmo 1 la estimación  $u_i^{(j)}$  corresponde a la última actualización del algoritmo ADMM y su formulación está dada por

$$u_i^{(j)} = u_i^{(j)} + s_i^{(j)} - \Phi_L \epsilon_i^{(j)} \quad (40)$$

La secuencia descrita anteriormente es consolidada en el algoritmo 1, en donde se presenta cada paso de actualización para el aprendizaje del diccionario espacio-espectral conjunto propuesto. Como entrada el algoritmo recibe una imagen HS  $x$  e inicializa de forma aleatoria el diccionario  $\Phi_L$  garantizado norma uno para cada

---

<sup>34</sup> Bradley Efron y col. «Least angle regression». En: *Annals of statistics* 32.2 (2004), págs. 407-499.

una de sus columnas. Luego de  $T$  iteraciones, el algoritmo obtiene el diccionario entrenado  $\Phi_L$  para representar de forma escasa la información espacial y espectral en dicha imagen.

---

**Algoritmo 1** Aprendizaje en línea del diccionario convolucional espacio espectral conjunto.

---

- 1: **Entrada:** Imagen espectral  $\mathbf{x}$ , Inicialización aleatoria del diccionario  $\Phi_L$ .
  - 2: División en subimágenes  $\mathbf{x}_i = \mathbf{F}_i \mathbf{x}$
  - 3: Inicialización  $\mathbf{s}_i^{(j)} = \frac{1}{n^2} \mathbf{D}_j \mathbf{x}_i$ ,  $\mathbf{u}_i^{(j)} = 0$
  - 4: **para** iteración = 1: T **hacer**
  - 5:     Actualización de la representación escasa local mediante LASSO:  

$$\boldsymbol{\varepsilon}_i^{(j)} = \underset{\hat{\boldsymbol{\varepsilon}}_i^{(j)}}{\operatorname{argmin}} \frac{\rho}{2} \|\mathbf{s}_i^{(j)} - \Phi_L \hat{\boldsymbol{\varepsilon}}_i^{(j)} + \mathbf{u}_i^{(j)}\|_2^2 + \lambda \|\hat{\boldsymbol{\varepsilon}}_i^{(j)}\|_1$$
  - 6:     Reconstrucción de las entidades  

$$\mathbf{p}_i^{(j)} = \frac{1}{\rho} \mathbf{D}_j \mathbf{x}_i + \Phi_L \boldsymbol{\varepsilon}_i^{(j)} - \mathbf{u}_i^{(j)}$$
  - 7:     Agregación de las entidades  

$$\hat{\mathbf{x}}_i = \sum_{j=1}^{N_d M_d L} \mathbf{D}_j^T \mathbf{p}_i^{(j)}$$
  - 8:     Actualización de  $\mathbf{s}_i^{(j)}$  usando el Laplaciano local:  

$$\mathbf{s}_i^{(j)} = \mathbf{p}_i^{(j)} - \frac{1}{\rho + n} \mathbf{D}_j \hat{\mathbf{x}}_i \quad \text{ver Ec. (38)}$$
  - 9:     Actualización de la variable dual:  

$$\mathbf{u}_i^{(j)} = \mathbf{u}_i^{(j)} + \mathbf{s}_i^{(j)} - \Phi_L \boldsymbol{\varepsilon}_i^{(j)} \quad \text{ver Ec. (40)}$$
  - 10:     Actualización del diccionario  $\Phi_L$   

$$\Phi_L, \boldsymbol{\varepsilon}_i^{(j)} = \underset{\Phi_L, \hat{\boldsymbol{\varepsilon}}_i^{(j)}}{\operatorname{argmin}} \sum_{i=1}^{N_s} \sum_{j=1}^{N_d M_d L} \frac{\rho}{2} \|\mathbf{s}_i^{(j)} - \Phi_L \hat{\boldsymbol{\varepsilon}}_i^{(j)} + \mathbf{u}_i^{(j)}\|_2^2$$

$$s.t. \operatorname{supp}(\boldsymbol{\varepsilon}_i^{(j)}) = \operatorname{supp}(\hat{\boldsymbol{\varepsilon}}_i^{(j)}) \quad \text{ver Ec. (39)}$$
  - 11: **fin para**
  - 12: **Salida:** Diccionario entrenado  $\Phi_L$ , coeficientes escasos  $\left\{ \left\{ \boldsymbol{\varepsilon}_i^{(j)} \right\}_{i=1}^{N_s} \right\}_{j=1}^{N_d M_d L}$  y entidades  $\left\{ \left\{ \mathbf{s}_i^{(j)} \right\}_{i=1}^{N_s} \right\}_{j=1}^{N_d M_d L}$ .
-

## 5. PROCESO DE RESTAURACIÓN DE IMÁGENES ESPECTRALES

En el capítulo anterior se estableció el algoritmo base propuesto para aprender en línea un diccionario convolucional espacio-espectral conjunto. En este caso, el entrenamiento del diccionario será evaluado en los escenarios de restauración de imágenes espectrales cuando fuentes de ruido, borrosidad o resolución insuficiente causan distorsiones en la imagen. En general, la tarea de restauración de imagen ya ha sido abordada bajo la teoría de representación escasa y en la actualidad el estado del arte corresponde a estructuras de diccionario que aprovechan la información espacial y espectral de la imagen en forma separable. Por lo tanto, esta tesis provee una estrategia que aprovecha la información espacial y espectral de forma conjunta para lograr abstraer la mayor cantidad de características intrínsecas de la imagen e incluso aquellas que se ignoran cuando se decide analizar de forma separable la información. Otra ventaja del método propuesto es la elección de una estrategia en línea para la fase de entrenamiento de tal forma que la única muestra que conoce el algoritmo en esta fase es la imagen espectral distorsionada, ignorando así la necesidad de otros algoritmos para obtener grandes conjuntos de datos de entrenamiento. Adicionalmente, el algoritmo propuesto realiza una partición de la imagen que denotamos división en subimágenes con el fin de evitar el alto costo de almacenamiento que implicaría considerar la imagen espectral  $N \times M \times L$  como la muestra de entrenamiento completa. Las dimensiones de estas subimágenes son seleccionadas estratégicamente para encontrar un balance entre los recursos de almacenamiento disponibles y para mantener la consistencia del sentido de globalidad.

Matemáticamente, la imagen distorsionada  $\mathcal{Y} \in \mathbb{R}^{N \times M \times L}$  es dividida en subimágenes  $\mathcal{Y}_i \in \mathbb{R}^{P \times P \times L}$  las cuales conforman el conjunto de imágenes de entrada del algoritmo. Este conjunto es dividido de forma aleatoria en 70 % de subimágenes para el conjunto de datos de entrenamiento y 30 % de subimágenes para el conjunto de

datos de evaluación. La división del conjunto de imágenes  $\mathcal{Y}_i$  es llevada a cabo para evaluar el diccionario propuesto en función de datos conocidos y desconocidos, lo que permite concluir de forma correcta el impacto real sobre su rendimiento para la tarea de restauración de imágenes espectrales. A partir del conjunto de datos de entrenamiento se emplea el algoritmo descrito en Algoritmo 1 para entrenar el diccionario  $\Phi_L$  bajo el enfoque local propuesto. Una vez, el diccionario entrenado es obtenido se emplea un algoritmo de codificación escasa para restaurar las subimágenes de evaluación de acuerdo a la siguiente ecuación  $\mathbf{y}_i = \sum_{j=1}^{N_d M_d L} \mathbf{D}_j^T \mathbf{s}_i^{(j)}$ . En la Figura 5 se muestra en detalle el procedimiento que se lleva a cabo para la restauración de imágenes, incluyendo entradas, salidas y su tratamiento para ejecutar el aprendizaje del diccionario espacio-espectral descrito en Algoritmo 1.

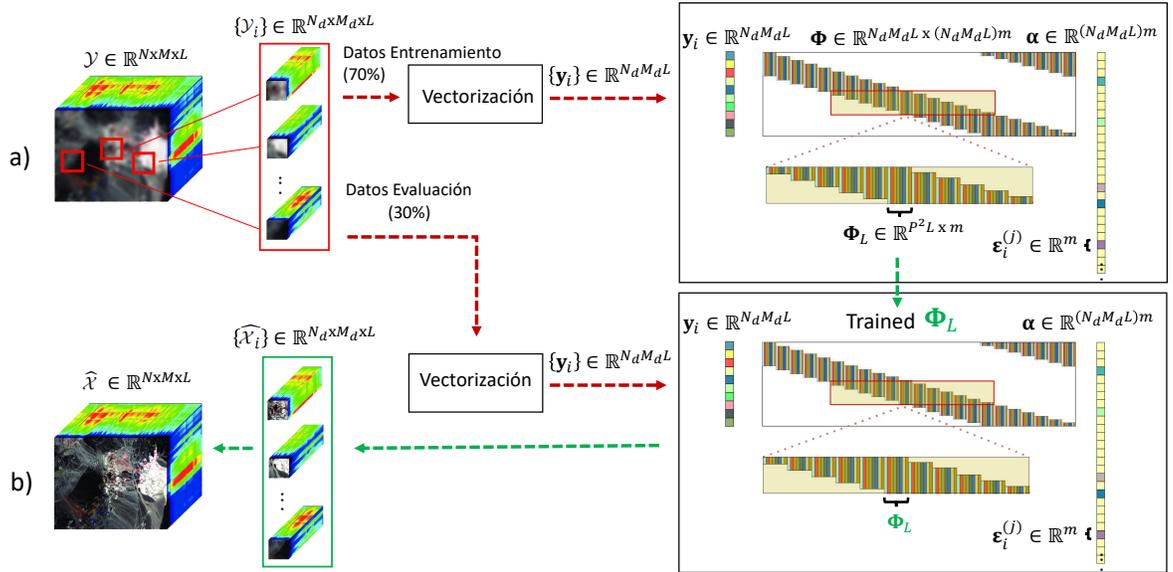


Figura 5. a) Entrenamiento del diccionario  $\Phi_L$ : La imagen espectral distorsionada  $\mathcal{Y}$  es dividida en subimágenes  $\mathcal{Y}_i \in \mathbb{R}^{N_d M_d L}$  para reducir la complejidad computacional. Posteriormente, el conjunto de subimágenes  $\mathcal{Y}_i$  es dividido en dos conjuntos de datos conformados por 70% de las muestras para entrenamiento y 30% de las muestras para evaluación. Así, los datos de entrenamiento son usados para estimar el diccionario espacio espectral  $\Phi_L$  bajo el algoritmo propuesto basado en localidad y los datos de validación son empleados para soportar la calidad de su estimación. b) Codificación de la representación escasa: Una vez la fase de entrenamiento a finalizado, el diccionario resultante  $\Phi_L$  es empleado en el algoritmo de codificación escasa para estimar los coeficientes  $\epsilon_i^{(j)}$  asociados al 30% de los datos de evaluación restantes, lo que es más conocido como la etapa de evaluación. En forma conjunta, los coeficientes  $\epsilon_i^{(j)}$  asociados a los datos de entrenamiento (70%) son también estimados para calcular la imagen espectral  $\hat{\mathcal{X}}$  sin distorsión.

Asumiendo que la imagen de entrada al algoritmo de restauración presenta alguna degradación o distorsión, se tiene que  $y = Hx$ , siendo  $H$  la matriz de degradación. De esta forma, se presenta la formulación matemática para cuando  $H$  modela la matriz de inclusión de ruido, inclusión de borrosidad y reducción de la resolución, respectivamente.

## 5.1. Eliminación de ruido

Las medidas  $y$  son una versión ruidosa de la imagen original y se modelan por la inclusión del término de ruido Gaussiano  $\mathbf{n}$ , tal que  $y = \mathbf{I}x + \mathbf{n}$  siendo la identidad su matriz de degradación  $\mathbf{H} = \mathbf{I}$ . El objetivo en el problema de eliminación de ruido es estimar la imagen desconocida  $x$  a partir de las medidas ruidosas  $y$ . De esta forma, la eliminación del ruido puede ser abordada bajo la minimización del siguiente problema de optimización

$$\alpha_i = \underset{\alpha_i}{\operatorname{argmin}} \|\mathbf{y}_i - \mathbf{I}\Phi\alpha_i\|_2 + \lambda\|\alpha_i\|_1, \text{ para } i = 1, 2, \dots, N_s, \quad (41)$$

donde  $\Phi$  representa el diccionario previamente entrenado y  $\lambda$  es el parámetro de regularización que controla el nivel de escasez. De acuerdo, al capítulo anterior se reescribe la descomposición escasa para cada subdivisión  $y_i$  en función de sus entidades, tal que el problema de minimización en la Ec. (41) es reformulado como

$$\begin{aligned} (\Phi_L, \mathbf{s}_i, \mathbf{u}_i) = & \underset{\Phi_L, \{\boldsymbol{\varepsilon}_i^{(j)}\}_{j=1}^{N_d M_d L}, \{\mathbf{s}_i^{(j)}\}_{j=1}^{N_d M_d L}, \{\mathbf{u}_i^{(j)}\}_{j=1}^{N_d M_d L}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y}_i - \mathbf{I} \sum_{j=1}^{N_d M_d L} \mathbf{D}_j^T \mathbf{s}_i^{(j)}\| \\ & + \sum_{j=1}^{N_d M_d L} \left( \lambda \|\boldsymbol{\varepsilon}_i^{(j)}\|_1 + \frac{\rho}{2} \|\mathbf{s}_i^{(j)} - \Phi_L \boldsymbol{\varepsilon}_i^{(j)} + \mathbf{u}_i^{(j)}\|_2^2 \right). \end{aligned} \quad (42)$$

Para el caso de eliminación de ruido, el problema de minimización en la Ec. (42) puede ser direccionado bajo el Algoritmo 1 siguiendo cada uno de sus pasos ya que la matriz de distorsión corresponde a la matriz identidad  $\mathbf{I}$  y no altera ningún paso de actualización.

## 5.2. Borrosidad

En el escenario de borrosidad las mediciones  $y$  son modeladas por una matriz de degradación  $\mathbf{H} = \mathbf{B}$ , tal que  $y = \mathbf{B}x$ . De igual forma que en los problemas de restauración de imagen se quiere recuperar la imagen original  $x$  a partir de las medidas con borrosidad  $y$ . Esta tarea es abordada bajo la formulación del siguiente problema de optimización

$$\alpha_i = \underset{\alpha_i}{\operatorname{argmin}} \quad \|y_i - \mathbf{B}\Phi\alpha_i\|_2 + \lambda\|\alpha_i\|_1. \quad (43)$$

En forma similar al caso de eliminación de ruido presentado en la sección anterior, la Ec. (43) puede ser reescrita como

$$\begin{aligned} (\Phi_L, \mathbf{s}_i, \mathbf{u}_i) = \underset{\{\boldsymbol{\varepsilon}_i^{(j)}\}_{j=1}^{N_d M_d L}, \{\mathbf{s}_i^{(j)}\}_{j=1}^{N_d M_d L}, \{\mathbf{u}_i^{(j)}\}_{j=1}^{N_d M_d L}}{\operatorname{argmin}} & \quad \frac{1}{2} \|y_i - \mathbf{B} \sum_{j=1}^{N_d M_d L} \mathbf{D}_j^T \mathbf{s}_i^{(j)}\| \\ & + \sum_{j=1}^{N_d M_d L} \left( \lambda \|\boldsymbol{\varepsilon}_i^{(j)}\|_1 + \frac{\rho}{2} \|\mathbf{s}_i^{(j)} - \Phi_L \boldsymbol{\varepsilon}_i^{(j)} + \mathbf{u}_i^{(j)}\|_2^2 \right). \end{aligned} \quad (44)$$

De acuerdo a la solución planteada en el capítulo anterior, la inclusión de la matriz de degradación  $\mathbf{B}$  afecta solo la estimación de las entidades  $\mathbf{s}_i^{(j)}$ . Por lo tanto, un nuevo paso de actualización es formulado para estimar cada una de las entidades  $\{\{\mathbf{s}_i^{(j)}\}_{j=1}^{N_d M_d L}\}_{i=1}^{N_s}$ , y se escribe como

$$\mathbf{s}_i^{(j)} = \left( \frac{1}{\rho} \mathbf{D}_j y_i + \Phi_L \boldsymbol{\varepsilon}_i^{(j)} - \mathbf{u}_i^{(j)} \right) - \mathbf{D}_i \left( \frac{1}{\rho + n} \mathbf{B} \sum_{j=1}^{N_d M_d L} \mathbf{D}_j^T \left( \frac{1}{\rho} \mathbf{D}_j y_i + \Phi_L \boldsymbol{\varepsilon}_i^{(j)} - \mathbf{u}_i^{(j)} \right) \right). \quad (45)$$

### 5.3. Super-resolución

Siguiendo un procedimiento similar al realizado en la sección 5.2 se tiene que para el caso de super-resolución cuando  $\mathbf{H} = \mathbf{PB}$ , la actualización de las entidades  $\{\{\mathbf{s}_i^{(j)}\}_{j=1}^{N_d M_d L}\}_{i=1}^{N_s}$  está dada por la siguiente ecuación

$$\mathbf{s}_i^{(j)} = \left( \frac{1}{\rho} \mathbf{D}_j \mathbf{y}_i + \Phi_L \boldsymbol{\varepsilon}_i^{(j)} - \mathbf{u}_i^{(j)} \right) - \mathbf{D}_i \left( \frac{1}{\rho + n} \mathbf{PB} \sum_{j=1}^{N_d M_d L} \mathbf{D}_j^T \left( \frac{1}{\rho} \mathbf{D}_j \mathbf{y}_i + \Phi_L \boldsymbol{\varepsilon}_i^{(j)} - \mathbf{u}_i^{(j)} \right) \right). \quad (46)$$

De forma general, observe que para ninguno de los casos de distorsión, la inclusión de una matriz de degradación  $\mathbf{H}$  afecta los pasos de actualización de las variables  $\Phi_L, \boldsymbol{\varepsilon}_i^{(j)}, \mathbf{u}_i^{(j)}$  y el procedimiento para la restauración de imagen puede realizarse siguiendo el Algoritmo 1.

## 6. EXPERIMENTOS NUMÉRICOS

En esta sección se realiza una evaluación y comparación del rendimiento de restauración de imágenes espectrales usando el diccionario convolucional espacio-espectral conjunto propuesto. Para ello, la imagen espectral distorsionada  $\mathcal{Y} \in \mathbb{R}^{N \times M \times L}$  es dividida para conformar el conjunto de muestra de datos de  $N_s$  subimágenes  $\mathcal{Y}_i \in \mathbb{R}^{N/8 \times N/8 \times L}$ , las cuales son posteriormente sometidas a un preprocesamiento para la limpieza de los datos. Posteriormente, el conjunto de  $N_s = 64$  muestras es dividido en 70 % de datos para entrenamiento ( $N_s * 0.7$ ) y 30 % de datos para evaluación ( $N_s * 0.3$ ). Una vez dividido el conjunto de datos, se realiza la extracción de parches  $\mathcal{Y}_i^{(j)} \in \mathbb{R}^{5 \times 5 \times L}$  o en su forma vectorial  $\mathbf{y}_i^{(j)} \in \mathbb{R}^{25L}$  para cada subdivisión de imagen  $\mathbf{y}_i$  y se realiza la inicialización de las variables  $\Phi_L$ ,  $\mathbf{s}_i^{(j)}$  y  $\mathbf{u}_i^{(j)}$  para comenzar la ejecución del algoritmo propuesto. Específicamente, el diccionario  $\Phi_L$  es inicializado de forma aleatoria garantizando norma uno en sus columnas, mientras que, las variables  $\mathbf{s}_i^{(j)}$  y  $\mathbf{u}_i^{(j)}$  son inicializadas en cero. Para la solución de los problemas LASSO y aprendizaje del diccionario en el Algoritmo 1 se usaron los métodos LARS y K-SVD, respectivamente. No obstante, cualquier otro algoritmo de los propuestos en el estado del arte puede ser usado para la solución de estos dos problemas aprovechando la escalabilidad del algoritmo propuesto. Finalmente, la restauración de la imagen espectral es llevada a cabo por el paso de codificación escasa usando como base el diccionario entrenado  $\Phi_L$ . De esta forma, cada representación  $\varepsilon_i^{(j)}$  es estimada a partir del algoritmo LASSO para obtener una estimación de la imagen espectral restaurada.

En los experimentos realizados se usaron dos conjuntos de datos del estado del arte: Chikusei y Washington DC MALL.

- Conjunto de datos Chikusei: Esta escena de estudio tiene un tamaño de  $540 \times$

540 en la dimensión espacial y 128 bandas espectrales en el rango  $363nm$  a  $1018\mu nm$ . El punto central de la escena se encuentra en las coordenadas: 36.294946 N, 140.00838 OE y tiene una resolución espacial de  $2,5m$ <sup>35</sup>.

- Conjunto de datos Washington DC MALL: Esta escena consta de 420300 píxeles y cubre el rango de  $401.29nm$  a  $2473.16nm$  con 191 bandas espectrales a una resolución espectral de aproximadamente 10 nm y una resolución espacial de 3 m. Para este experimento se realizó una selección previa de 120 bandas espectrales, de este modo el conjunto de datos resultante tiene las dimensiones  $420 \times 300 \times 120$ .

Entre otros parámetros importantes se define el número de filtros del diccionario convolucional  $m = 2000$ , el tamaño de filtro  $5 \times 5 \times 191$ , el número máximo de valores diferentes de cero por entidad establecido en 90 y el parámetro de regularización asociado al Lagrangiano  $\lambda = 0.2$ . El rendimiento del método fue demostrado mediante métricas que evalúan la calidad de reconstrucción espacial PSNR, la similitud estructural (SSIM), la similitud del espectro (SAM), la calidad espectral general (ERGAS), la similitud de características de imagen (FSIM), y la distorsión estructural (UIQI).

## 6.1. RESULTADOS NUMERICOS

A continuación, se presentan los resultados numéricos obtenidos para los casos de eliminación de ruido, borrosidad y super resolución usando el diccionario propuesto en los dos conjuntos de datos de prueba, para cada caso. Con el objetivo de calcular

---

<sup>35</sup> Hayder Hasan, Helmi ZM Shafri y Mohammed Habshi. «A comparison between support vector machine (SVM) and convolutional neural network (CNN) models for hyperspectral image classification». En: *IOP Conference Series: Earth and Environmental Science*. Vol. 357. 1. IOP Publishing. 2019, pág. 012035.

el rendimiento del método propuesto y compararlo con el estado del arte, se realizaron comparaciones exhaustivas respecto a métodos de la literatura con respecto a cada tipo de distorsión. Para ello, las métricas de evaluación fueron seleccionadas de la siguiente forma, super resolución empleo las métricas PSNR, SSIM, FSIM, ERGAS y SAM, eliminación de ruido empleo las métricas PSNR, SSIM, ERGAS y SAM y reducción de borrosidad empleo las métricas PSNR, ERGAS, SAM y UIQI.

**6.1.1. Eliminación de ruido** La imagen distorsionada y es simulada agregando ruido gaussiano  $n$  de media cero y varianza 0.15 a todas las bandas espectrales de la imagen original  $x$ , En las tablas 1 y 2 se muestran los valores cuantitativos de las métricas de rendimiento para los diferentes métodos competitivos que solucionan el caso de eliminación de ruido

Tabla 1. Rendimiento de la imagen ruidosa, los métodos LRMR , LRTV, SSTV, LSSTV, LRTDTV, LRTDGS y el enfoque propuesto sobre el conjunto de datos Chi-kusei: PSNR, SSIM, FSIM, ERGAS y SAM.

Metrics	Ruido	LRMR <sup>36</sup>	LRTV <sup>37</sup>	SSTV <sup>38</sup>	LSSTV <sup>39</sup>	LRTDTV <sup>40</sup>	LRTDGS <sup>41</sup>	Propuesta
PSNR	16.48	32.14	33.23	29.34	31.49	33.46	36.96	<b>38.39</b>
SSIM	0.28	0.92	0.93	0.83	0.89	0.93	0.94	<b>0.94</b>
FSIM	0.63	0.95	0.96	0.92	0.94	0.96	0.96	<b>0.97</b>
ERGAS	633.77	101.80	90.27	145.80	113.07	86.81	82.75	<b>75.26</b>
SAM	0.73	0.16	0.14	0.24	0.19	<b>0.11</b>	0.12	0.16

Tabla 2. Rendimiento de la imagen ruidosa, los métodos LRMR, LRTV, SSTV, LSSTV, LRTDTV, LRTDGS y el enfoque propuesto sobre el conjunto de datos Washington DC Mall: PSNR, SSIM, FSIM, ERGAS y SAM.

Metrics	Ruido	LRMR	LRTV	SSTV	LSSTV	LRTDTV	LRTDGS	Propuesta
PSNR	16.42	33.09	36.32	31.18	31.76	37.88	40.01	<b>41.11</b>
SSIM	0.26	0.88	0.98	0.80	0.82	0.98	0.99	<b>0.99</b>
FSIM	0.38	0.88	0.97	0.81	0.83	0.97	<b>0.99</b>	0.98
ERGAS	351.06	52.89	37.27	68.31	64.72	31.49	24.60	<b>20.02</b>
SAM	0.29	0.03	0.02	0.05	0.04	0.02	0.02	<b>0.01</b>

**6.1.2. Borrosidad** La imagen distorsionada  $y$  es simulada agregando un efecto de borrosidad  $B$  mediante la operación de convolución entre cada banda espectral de la imagen original  $x$  y un núcleo Gaussiano con parámetros  $\sigma = 7$  y tamaño de dimensión  $8 \times 8$ . En las tablas 3 y 4 se muestran los valores cuantitativos de las métricas de rendimiento para los diferentes métodos competitivos que solucionan el caso de reducción de borrosidad.

Tabla 3. Rendimiento de los métodos HL, FPD, SSTV, DB-WLRTR y el enfoque propuesto sobre el conjunto de datos Chikusei: PSNR, SSIM, ERGAS y SAM.

Metrics	HL <sup>42</sup>	FPD <sup>43</sup>	SSTV <b>SSTV</b>	DB-WLRTR <sup>44</sup>	Propuesta
PSNR	32.59	33.16	33.08	<b>49.42</b>	48.92
SSIM	0.88	0.91	0.89	<b>0.99</b>	0.97
ERGAS	137.14	125.11	129.84	<b>20.87</b>	23.06
SAM	0.11	0.12	0.09	0.04	<b>0.03</b>

Tabla 4. Rendimiento de los métodos HL, FPD, SSTV, DB-WLRTR y el enfoque propuesto sobre el conjunto de datos Washington DC MALL: PSNR, SSIM, ERGAS y SAM.

Metrics	HL	FPD	SSTV	DB-WLRTR	Propuesta
PSNR	33.59	34.16	36.50	50.02	<b>52.7</b>
SSIM	0.88	0.90	0.89	0.99	<b>0.99</b>
ERGAS	135.30	125.67	131.82	15.21	<b>13.97</b>
SAM	0.21	0.16	0.11	0.03	<b>0.01</b>

**6.1.3. Super-resolucion** La distorsión PB sobre  $y$  es simulada primero agregando un efecto de borrosidad con un núcleo Gaussiano de tamaño  $8 \times 8$  y  $\sigma = 5$  y segundo aplicando una reducción de la resolución con un factor de escalamiento  $s = 6$  y  $s = 5$  para los conjuntos de datos Chikusei y Washington DC MALL, respectivamente. En las tablas 7 y 6 se muestran los valores cuantitativos de las métricas de rendimiento para los diferentes métodos competitivos que solucionan el caso de super resolución.

Tabla 5. Rendimiento de los métodos Bicubico, Bilineal, A2SDLHSI, HySure, NPTSR y el enfoque propuesto sobre el conjunto de datos Chikusei: PSNR, ERGAS, SAM y UIQI.

Metrics	Bicubico	Bilineal	A2SDLHSI <sup>45</sup>	HySure <sup>46</sup>	NPTSR <sup>47</sup>	Propuesta
PSNR	31.98	30.80	33.17	33.22	34.68	<b>36.09</b>
ERGAS	0.70	0.75	0.69	0.67	0.64	<b>0.61</b>
SAM	<b>2.85</b>	2.91	3.02	2.98	3.12	2.86
UIQI	0.94	0.94	0.95	0.96	0.95	<b>0.94</b>

Tabla 6. Rendimiento de los métodos Bicubico, Bilineal, A2SDLHSI, HySure, NPTSR y el enfoque propuesto sobre el conjunto de datos Washington DC MALL: PSNR, ERGAS, SAM y UIQI.

Metrics	Bicubico	Bilineal	A2SDLHSI	HySure	NPTSR	Propuesta
PSNR	31.99	31.01	34.02	34.12	35.96	<b>38.12</b>
ERGAS	0.65	0.72	0.65	0.62	0.60	<b>0.56</b>
SAM	<b>2.75</b>	2.83	2.91	2.87	3.06	2.72
UIQI	0.92	0.93	0.95	0.97	0.91	<b>0.91</b>

## 6.2. COMPLEJIDAD COMPUTACIONAL

En esta sección se reporta el tiempo de ejecución para el escenario de super-resolución el cual demostró los mayores tiempos para alcanzar la convergencia a la solución por parte del método propuesto. A continuación, se presenta la tabla resumen en donde se registra el tiempo de ejecución hasta convergencia que toma el método propuesto y los métodos del estado del arte A2SDLHSI, HySure y NPTSR para restaurar la imagen hiperespectral de baja resolución

Tabla 7. Tiempo de ejecución en segundos (s) para los métodos Bicubico, Bilineal, A2SDLHSI, HySure, NPTSR y el enfoque propuesto sobre los conjuntos de datos Chikusei y Washington DC MALL.

Datos / métodos	Bicubico	Bilineal	A2SDLHSI <sup>48</sup>	HySure <sup>49</sup>	NPTSR <sup>50</sup>	Propuesta
Chikusei	11.24	7.69	557.36	86.11	612.75	<b>2995.98</b>
Washington DC MALL	9.72	6.82	526.91	83.94	598.23	<b>2987.33</b>

Todo los experimentos fueron ejecutados sobre un Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz 3.41 GHz bajo el entorno de Matlab R2020b. Los valores registrados

en la tabla para los conjuntos de datos Chikusei y Washington DC MALL corresponden a un cálculo promedio del tiempo para diez ejecuciones diferentes. Como se puede observar en la tabla el método propuesto muestra el costo computacional más alto, sin embargo, su diferencia en tiempo es aceptable debido a la ganancia en contraparte que se obtiene en calidad de restauración basado en las características intrínsecas que se extraen del espacio no explorado y alto en dimensionalidad espacio-espectral conjunto. Adicionalmente, el rendimiento del algoritmo propuesto puede acelerarse con el uso de *hardware* GPU ya que la restauración de cada subimagen  $x_i$  puede ser realizada en paralelo.

### 6.3. RESTAURACION DE IMAGENES

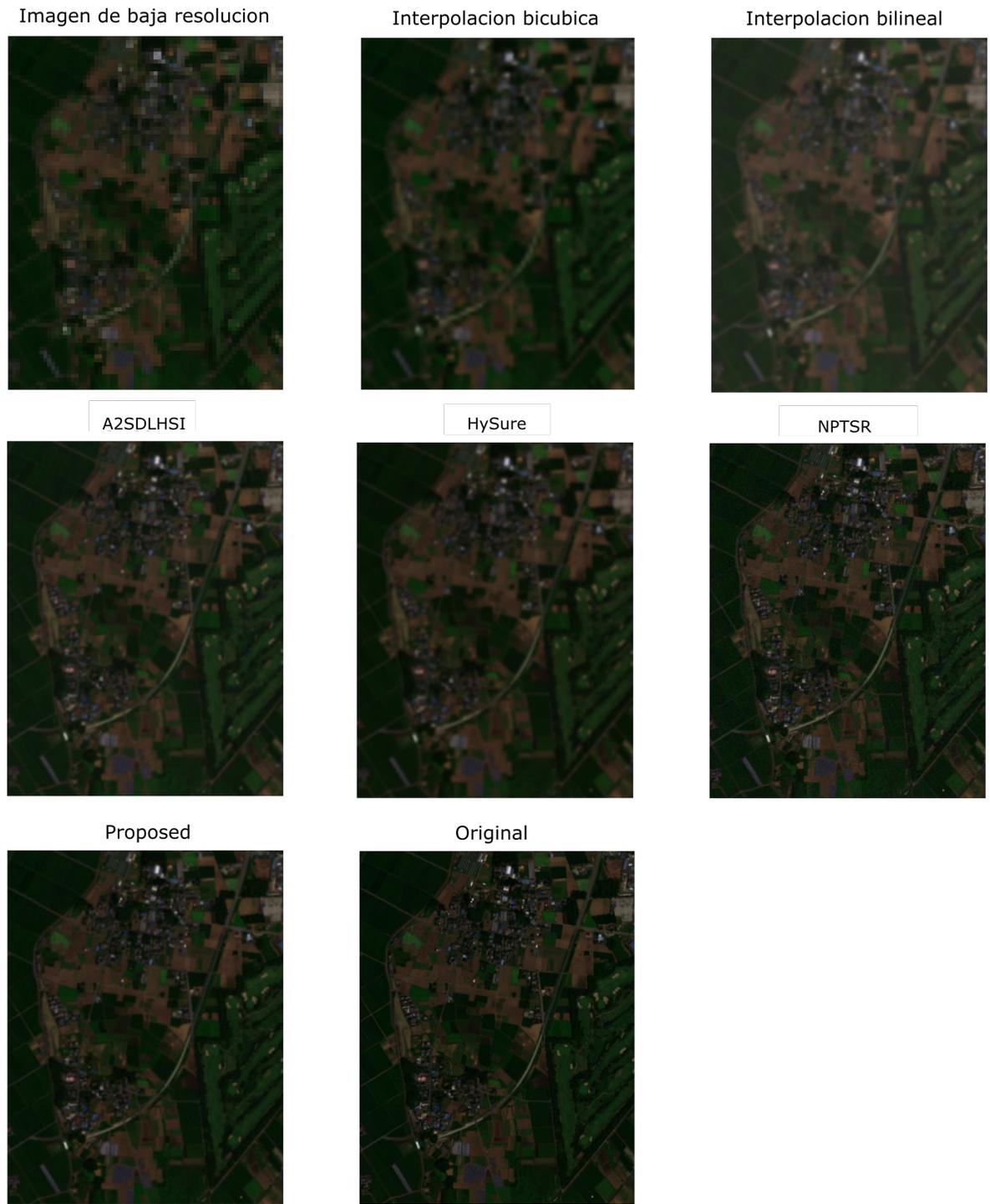


Figura 6. Comparación visual de los métodos bicubico, bilineal, A2SDLHSI, HySure, NPTSR y la propuesta para Super resolución usando el conjunto de datos Chikusei.

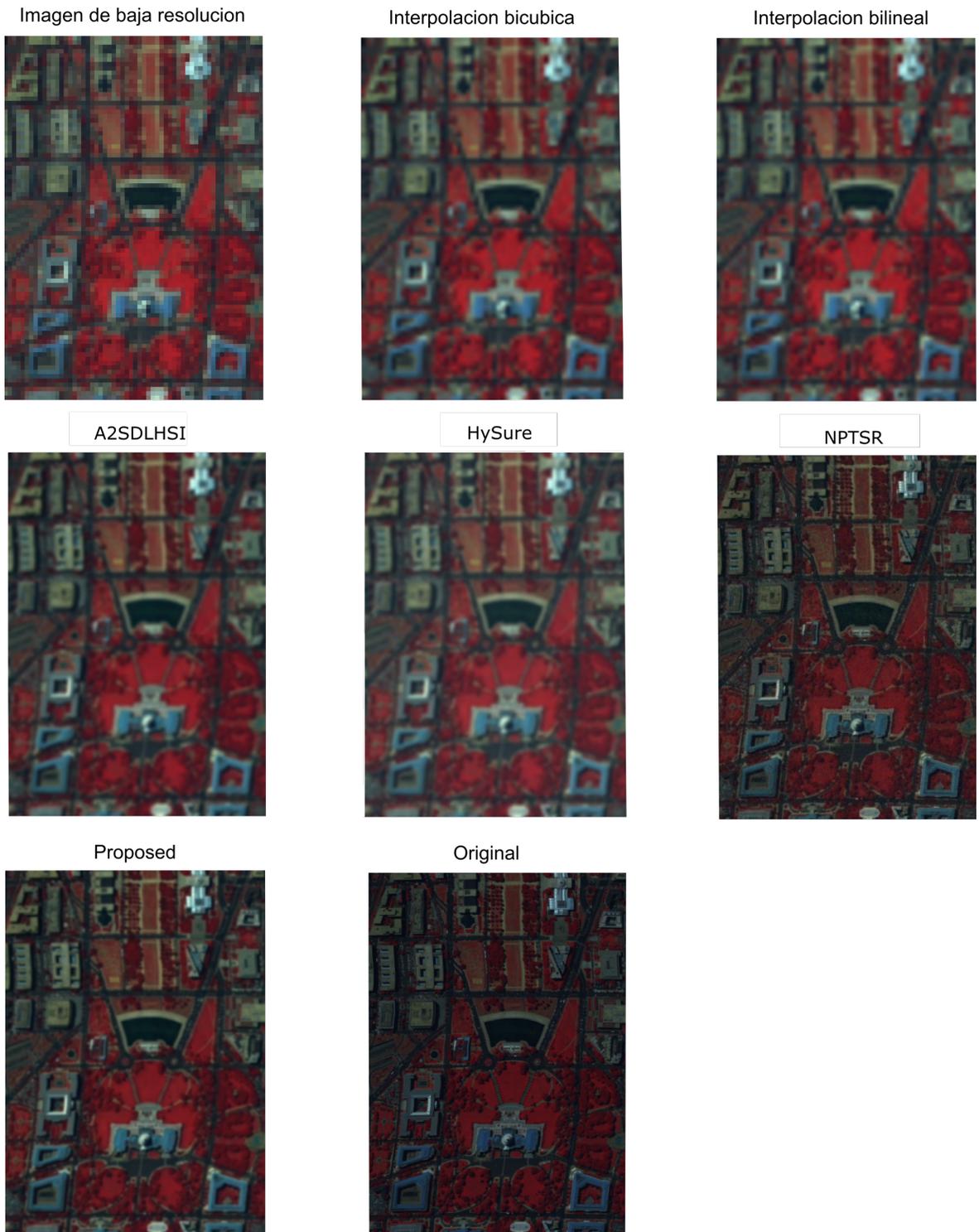


Figura 7. Comparación visual de los métodos bicubico, bilineal, A2SDLHSI, HySure, NPTSR y la propuesta para Super resolución usando el conjunto de datos Washington DC MALL.

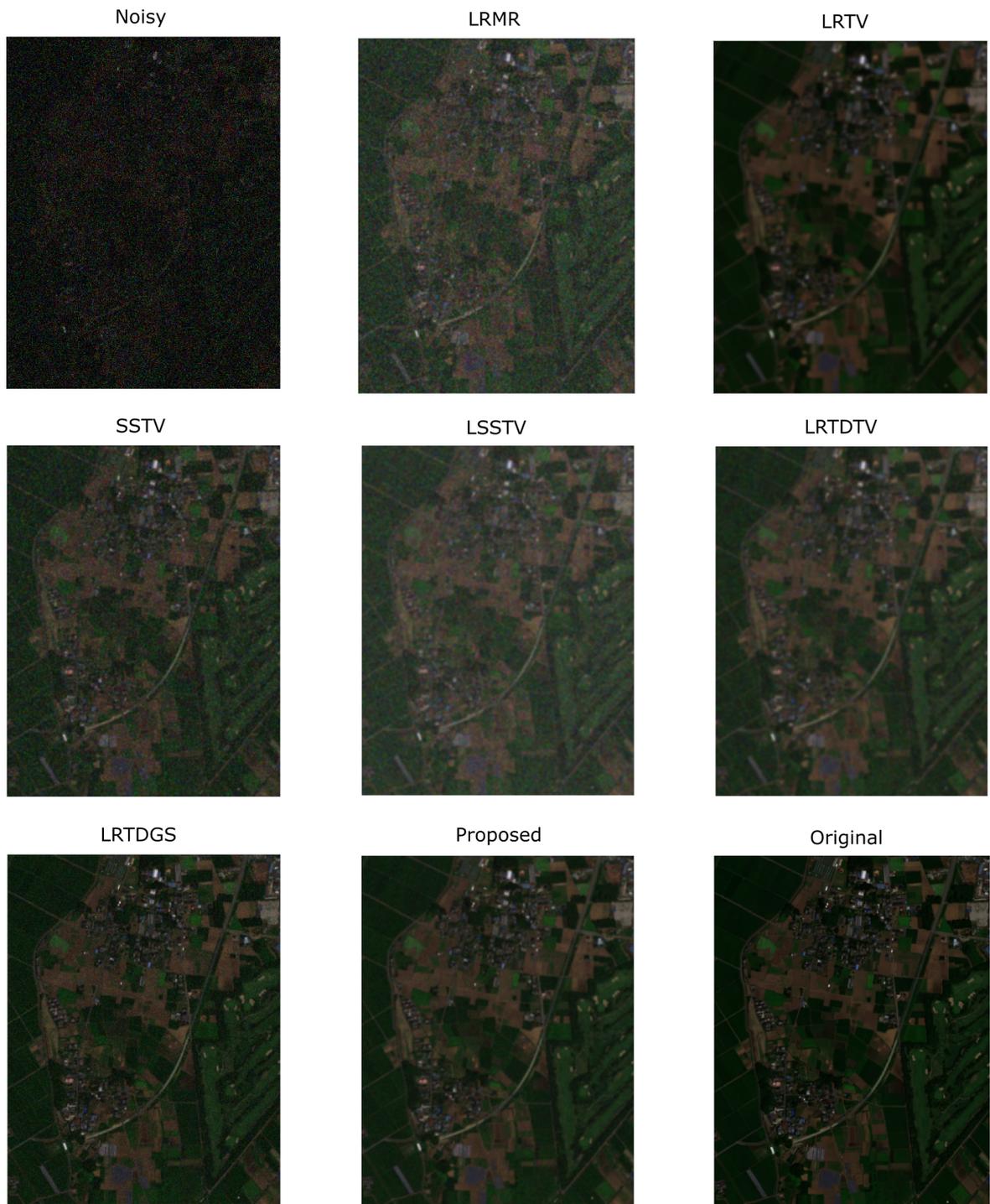


Figura 8. Comparación visual de los métodos LRMR, LRTV, SSTV, LSSTV, LRTDTV, LRTDGS y la propuesta para la eliminación de ruido usando el conjunto de datos Chikusei.



Figura 9. Comparación visual de los métodos LRMR, LRTV, SSTV, LSSTV, LRTDTV, LRTDGS y la propuesta para la eliminación de ruido usando el conjunto de datos Washington DC MALL.

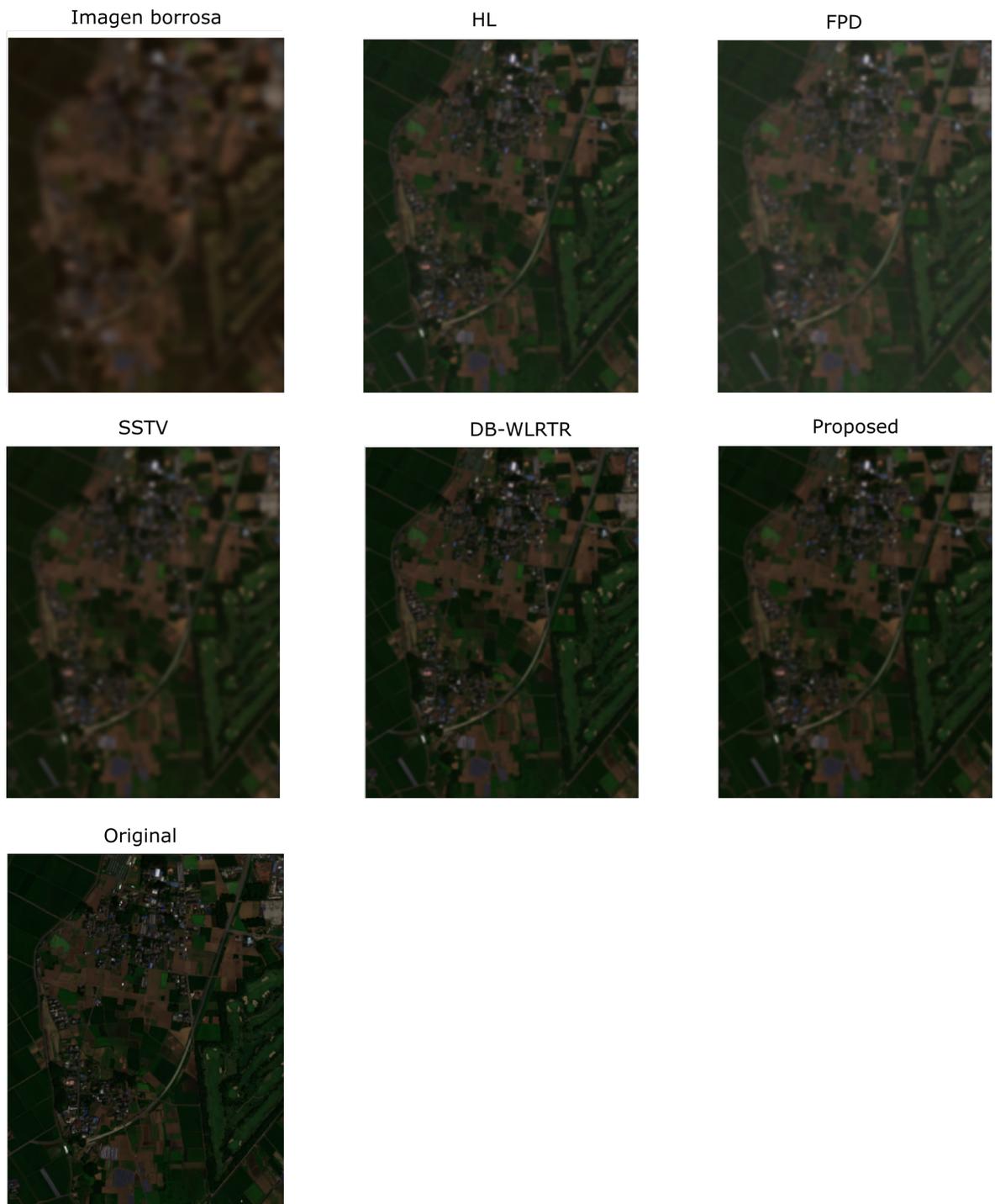


Figura 10. Comparación visual de los métodos HL, FPD, SSTV, DB-WLRTR, y la propuesta para la reducción de la borrosidad usando el conjunto de datos Chikusei.



Figura 11. Comparación visual de los métodos HL, FPD, SSTV, DB-WLRTR, y la propuesta para la reducción de la borrosidad usando el conjunto de datos Washington DC MALL.



## 7. CONCLUSIONES

En este trabajo se plantea una estructura de diccionario 3D que permite descubrir relaciones no exploradas en el espacio espacio-espectral de las imágenes espectrales. Adicionalmente, el planteamiento del diccionario es realizado para mitigar la carga computacional que implica abordar el entrenamiento desde un enfoque conjunto a partir de información espacial y espectral. En base a lo anterior, se plantea un diccionario convolucional basado en un enfoque local, el cual permite operar de forma local sobre información espacio-espectral y sin perder el sentido de globalidad. Por otra parte, el descubrimiento de nuevas características intrínsecas presentes en el espacio espacio-espectral se ve reflejado en una ganancia de las métricas de restauración de imagen con respecto a los métodos del estado del arte para los casos de super resolución, eliminación del ruido y reducción de borrosidad. Por otra parte, el algoritmo planteado puede ser fácilmente escalado para solucionar cualquier otro tipo de problema de restauración de imagen e incluso escalable para usar cualquier otro método en sus pasos de actualización.

## BIBLIOGRAFÍA

- Aggarwal, Hemant Kumar y Angshul Majumdar. «Hyperspectral image denoising using spatio-spectral total variation». En: *IEEE Geoscience and Remote Sensing Letters* 13.3 (2016), págs. 442-446 (vid. pág. 57).
- Aharon, Michal, Michael Elad y Alfred Bruckstein. «K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation». En: *IEEE Transactions on signal processing* 54.11 (2006), págs. 4311-4322 (vid. pág. 15).
- Arias, Kevin, Edwin Vargas y Henry Arguello. «Hyperspectral and Multispectral Image Fusion based on a Non-locally Centralized Sparse Model and Adaptive Spatial-Spectral Dictionaries». En: *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE. 2019, págs. 1-5 (vid. págs. 15, 16, 25, 26, 34).
- Barajas-Solano, C y col. «Tridimensional Convolutional Sparse Coding of Spectral Images». En: *Hyperspectral Imaging and Sounding of the Environment*. Optical Society of America. 2019, HTu3B-5 (vid. pág. 32).
- Bioucas-Dias, José M y col. «Hyperspectral remote sensing data analysis and future challenges». En: *IEEE Geoscience and remote sensing magazine* 1.2 (2013), págs. 6-36 (vid. pág. 13).
- Candès, Emmanuel J y Michael B Wakin. «An introduction to compressive sampling». En: *IEEE signal processing magazine* 25.2 (2008), págs. 21-30 (vid. pág. 14).
- Chang, Chein-I. *Hyperspectral data exploitation: theory and applications*. John Wiley & Sons, 2007 (vid. pág. 13).

- Chang, Yi y col. «Weighted low-rank tensor recovery for hyperspectral image restoration». En: *IEEE transactions on cybernetics* 50.11 (2020), págs. 4558-4572 (vid. pág. 58).
- Chen, Yong y col. «Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition». En: *IEEE transactions on cybernetics* (2019) (vid. págs. 14, 57).
- Cheng, Jian y col. «Remote sensing image fusion via wavelet transform and sparse representation». En: *ISPRS journal of photogrammetry and remote sensing* 104 (2015), págs. 158-173 (vid. pág. 15).
- Dian, Renwei, Leyuan Fang y Shutao Li. «Hyperspectral image super-resolution via non-local sparse tensor factorization». En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, págs. 5344-5353 (vid. pág. 27).
- Dong, Ganggang y col. «Target recognition in SAR images via sparse representation in the frequency domain». En: *Pattern Recognition* 96 (2019), pág. 106972 (vid. pág. 15).
- Dumitrescu, Bogdan y Paul Irofti. *Dictionary learning algorithms and applications*. Springer, 2018 (vid. pág. 16).
- Efron, Bradley y col. «Least angle regression». En: *Annals of statistics* 32.2 (2004), págs. 407-499 (vid. pág. 47).
- Fu, Ying y col. «Adaptive spatial-spectral dictionary learning for hyperspectral image denoising». En: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, págs. 343-351 (vid. pág. 26).

- Fu, Ying y col. «Adaptive spatial-spectral dictionary learning for hyperspectral image restoration». En: *International Journal of Computer Vision* 122.2 (2017), págs. 228-245 (vid. págs. 14, 16, 17, 25, 34, 59).
- Garcia-Cardona, Cristina y Brendt Wohlberg. «Convolutional dictionary learning». En: *arXiv preprint arXiv:1709.02893* (2017) (vid. pág. 16).
- «Convolutional dictionary learning: A comparative review and new algorithms». En: *IEEE Transactions on Computational Imaging* 4.3 (2018), págs. 366-381 (vid. pág. 38).
- «Convolutional dictionary learning for multi-channel signals». En: *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE. 2018, págs. 335-342 (vid. pág. 30).
- Hasan, Hayder, Helmi ZM Shafri y Mohammed Habshi. «A comparison between support vector machine (SVM) and convolutional neural network (CNN) models for hyperspectral image classification». En: *IOP Conference Series: Earth and Environmental Science*. Vol. 357. 1. IOP Publishing. 2019, pág. 012035 (vid. pág. 56).
- He, Wei y col. «Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration». En: *IEEE transactions on geoscience and remote sensing* 54.1 (2015), págs. 178-188 (vid. pág. 57).
- Henrot, Simon, Charles Soussen y David Brie. «Fast positive deconvolution of hyperspectral images». En: *IEEE Transactions on Image Processing* 22.2 (2012), págs. 828-833 (vid. pág. 58).

- Krishnan, Dilip y Rob Fergus. «Fast image deconvolution using hyper-Laplacian priors». En: *Advances in neural information processing systems* 22 (2009), págs. 1033-1041 (vid. pág. 58).
- Kumar, BK Shreyamsha, MNS Swamy y M Omair Ahmad. «Visual tracking using structural local DCT sparse appearance model with occlusion detection». En: *Multimedia Tools and Applications* 78.6 (2019), págs. 7243-7266 (vid. pág. 15).
- Li, Shutao y col. «Fusing hyperspectral and multispectral images via coupled sparse tensor factorization». En: *IEEE Transactions on Image Processing* 27.8 (2018), págs. 4118-4130 (vid. pág. 30).
- Lin, Xing y col. «Spatial-spectral encoded compressive hyperspectral imaging». En: *ACM Transactions on Graphics (TOG)* 33.6 (2014), págs. 1-11 (vid. págs. 17, 33).
- Panasyuk, Svetlana V y col. «Medical hyperspectral imaging to facilitate residual tumor identification during surgery». En: *Cancer biology & therapy* 6.3 (2007), págs. 439-446 (vid. pág. 13).
- Papayan, Vardan, Jeremias Sulam y Michael Elad. «Working locally thinking globally-part II: Stability and algorithms for convolutional sparse coding». En: *arXiv preprint arXiv:1607.02009* (2016) (vid. pág. 40).
- «Working locally thinking globally: Theoretical guarantees for convolutional sparse coding». En: *IEEE Transactions on Signal Processing* 65.21 (2017), págs. 5687-5701 (vid. pág. 40).
- Papayan, Vardan y col. «Convolutional dictionary learning via local processing». En: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, págs. 5296-5304 (vid. págs. 16, 38, 40).

- Sadeghi, Mostafa, Massoud Babaie-Zadeh y Christian Jutten. «Dictionary learning for sparse representation: A novel approach». En: *IEEE Signal Processing Letters* 20.12 (2013), págs. 1195-1198 (vid. pág. 15).
- Schaepman, Michael E y col. «Earth system science related imaging spectroscopy: An assessment». En: *Remote Sensing of Environment* 113 (2009), S123-S137 (vid. pág. 13).
- Schreiber, NM Förster y col. «The SINS survey: SINFONI integral field spectroscopy of z 2 star-forming galaxies». En: *The Astrophysical Journal* 706.2 (2009), pág. 1364 (vid. pág. 13).
- Shaw, Gary A y Hsiaohua K Burke. «Spectral imaging for remote sensing». En: *Lincoln laboratory journal* 14.1 (2003), págs. 3-28 (vid. pág. 34).
- Simoës, Miguel y col. «A convex formulation for hyperspectral image superresolution via subspace-based regularization». En: *IEEE Transactions on Geoscience and Remote Sensing* 53.6 (2014), págs. 3373-3388 (vid. pág. 59).
- Song, Xiaorui, Lingda Wu y Hongxing Hao. «Hyperspectral image denoising based on adaptive sparse representation». En: *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*. IEEE. 2018, págs. 735-739 (vid. pág. 15).
- Wang, Qiang y col. «Low rank constraint and spatial spectral total variation for hyperspectral image mixed denoising». En: *Signal Processing* 142 (2018), págs. 11-26 (vid. pág. 57).
- Wang, Yao y col. «Hyperspectral image restoration via total variation regularized low-rank tensor decomposition». En: *IEEE Journal of Selected Topics in Ap-*

*plied Earth Observations and Remote Sensing* 11.4 (2017), págs. 1227-1243 (vid. pág. 57).

Wang, Yu Winston y col. «Multiplexed optical imaging of tumor-directed nanoparticles: a review of imaging systems and approaches». En: *Nanotheranostics* 1.4 (2017), pág. 369 (vid. pág. 34).

Xu, Yang y col. «Nonlocal patch tensor sparse representation for hyperspectral image super-resolution». En: *IEEE Transactions on Image Processing* 28.6 (2019), págs. 3034-3047 (vid. págs. 14, 30, 59).

Yin, Haitao, Shutao Li y Leyuan Fang. «Simultaneous image fusion and super-resolution using sparse representation». En: *Information Fusion* 14.3 (2013), págs. 229-240 (vid. pág. 22).

Zha, Zhiyuan y col. «Image restoration using joint patch-group-based sparse representation». En: *IEEE Transactions on Image Processing* 29 (2020), págs. 7735-7750 (vid. pág. 14).

Zhang, Hongyan y col. «Hyperspectral image restoration using low-rank matrix recovery». En: *IEEE transactions on geoscience and remote sensing* 52.8 (2013), págs. 4729-4743 (vid. págs. 14, 57).