Optical Solutions for Spectral Imaging Inverse Problems with a Shift-Variant System.

Sergio Urrea[†], Roman Jacome[†], M. Salman Asif[‡], Henry Arguello[†] and Hans Garcia[†]
[†] Universidad Industrial de Santander
[‡] University of California Riverside
{sergio2228328@correo,roman2162474@correo,henarfu@,hans.garcia@saber}.uis.edu.co
sasif@ucr.edu

Abstract

Inverse problems in spectral imaging have been addressed in the state-of-the-art by encoding scenes to alleviate the ill-posedness, leveraging the knowledge of the forward model of the system. Recent studies have demonstrated that optimizing these coding elements improves the performance of solving the inverse problem. Specifically, to include a coding element without sacrificing the light throughput of the optical system, Diffractive Optical Elements (DOEs) have been employed. Recent works have highlighted the significance of shift-variant systems, which allows an optimized coding for each spatial portion of the scene and for each wavelength. With this in mind, this work proposes a shift-variant optical system using double-phase coding elements by implementing a double DOE architecture. The results show that using this proposed double-DOE architecture leads to better results in terms of high-level tasks, such as spectral image reconstruction and spatial-spectral super-resolution. Additionally, this work proves the shift-variant nature of the double-DOE architecture.

1. Introduction

Spectral imaging (SI) combines synergistic spectroscopy and imaging techniques to acquire high-quality spatial and spectral information, yielding in a 3D data cube. While a wide range of research fields have harnessed this type of data in various applications such as in remote sensing [13], medical imaging [21], and surveillance [7] among others, the acquisition of this type of information is still an open problem. Traditional SI is based on scanning portions of the spatio-spectral scene in several sensor integration times [14]. This scanning-based approach yields in highly time-consumption acquisition, limiting their use in dynamic scenes [5].

To overcome this limitation, compressive spectral imaging (CSI) methods have been proposed to capture the entire spatio-spectral cube in single-coded measurements [2, 30]. These methods rely on employing optical coding elements to modulate the spatial-spectral information and employ a 2D focal plane array (FPA) sensor to integrate the coded incident light. Consequently, an inverse problem is solved to recover the spectral image. Additionally, several inherent spectral imaging can be solved by using compressed measurements, such as super-resolution [29], and fusion [27] among others. Here, the quality of the estimate strongly depends on both the optical coding system and the decoding algorithms. Tremendous efforts have been made to develop highly precise reconstruction. For this task, several approaches have been proposed, spanning from traditional model-based convex optimization-based algorithms [23, 10] to black-box deep learning-based approaches [26, 25] and lastly hybrid model-based data-driven methods in the form of the so-called unrolling networks [17, 16]. These last methods are the current state-of-the-art in CSI recovery.

Complementary to algorithmic development, the acquisition system design has also been of great interest. First CSI optical architectures were based on coded apertures (CA) to modulate the spatial information and dispersive elements to codify the spectral elements. Some examples of these types of systems are the coded aperture snapshot spectral imager (CASSI) [28] and dual dispersive architecture [22] among others. These approaches employ amplitude codification via the CA which suffers from loss of incident energy in the sensor producing a lower signal-to-noise ratio. Additionally, the optical architectures are very complex and huge which difficult its implementation. Lately, phase codification has been explored for CSI, in which a diffractive optical element (DOE) is employed to modulate both spatial and spectral information, some examples of the use of DOEs employ patterns to design the height map, as in [18] where the proposed approach achieves a PSF that rotate for each wavelength, and in [6] is used deep neural

networks to learn the height-maps jointly with the reconstruction network, to achieve depth-dependent PSF to invert hyperspectral-depth problems. These systems are modeled as the convolution between the underlying spectral image by a wavelength-variant point spread function (PSF), determined by the DOE structure. Thus, these systems overcome the energy loss issue and also are very compact. Usually, the DOE's PSF is spatial shift-invariant increasing the ill-posedness of the inverse problem. Then, in [4] a shift-variant DOE-based optical system is proposed by including a colored CA (CCA) between the DOE and the sensor. However, the CCA also reduces the incident light energy.

Therefore, to promote a shift-variant optical system for CSI, a double DOE codification optical architecture is proposed. We prove that the PSF produced by this system varies along the spatial and spectral dimensions, yielding a richer codification. The two DOEs heightmaps are optimized by employing an End-to-End (E2E) optimization approach [3]. Here the DOEs are parametrized using Zernike polynomials [24] and we optimize the coefficients of each polynomial. These coefficients are jointly optimized with the weights of the deep neural network that performs the reconstruction of the spectral image as shown in Fig. 2(a). The employed reconstruction network is based on the unrolling algorithm, which induces a better optimization of the sensing systems since this network depends on the forward sensing operator. Two recovery problems are evaluated by using our approach, the first one is recovering high-fidelity spectral image from the coded measurements and the second is a spatial-super resolution problem where the FPA is assumed to have a high pitch size than the underlying spectral image. We compared the proposed optical system with a single-DOE-based system (shift-invariant) also trained in an E2E manner, and a non-data-driven DOE design named Spiral [18]. We show that our proposed design yields better reconstruction performance than shift-invariant approaches.

2. Mathematical Model of Double-DOE Architecture

This section describes the mathematical model of the double-DOE optical architecture. Fig. 1 shows an illustration of the proposed system, which is mainly comprised of two DOEs and an RGB sensor where for two different light point sources, the response of the system is spatially different resulting in a shift-variant response. The light propagation of DOE-based architectures can be modeled using mainly three operations; the codification of the wavefront phase by a DOE, the propagation between optical elements, and the sensing process given by the RGB sensor. For this work, the light propagation between optical elements is modeled via Fresnel propagation [12]. Then, for a incident wavefront $u(x, y, \lambda)$ Fresnel propagation operator for a distance z is defined as

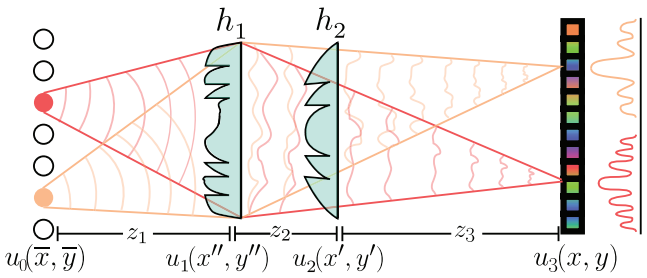

Figure 1. Illustration of the proposed double-DOE architecture exhibiting shift-variant property where for two light sources, the system produces different response.

$$\mathcal{P}\{u(x, y, \lambda), z\} = \frac{e^{jkz}}{jz\lambda} \iint u(x, y, \lambda) e^{\frac{jk}{2z}((x'-\bar{x})^2+(y'-y)^2)}\, dx\, dy. \quad (1)$$

The Eq. (1) can be rewritten using the convolution operator as

$$\mathcal{P}\{u(x, y, \lambda), z\} = \frac{e^{jkz}}{jz\lambda} \left(u(x, y, \lambda) * e^{\frac{jk}{2z}(x^2+y^2)} \right). \quad (2)$$

Then, the effect of a DOE, with a height map $h(x'', y'')$ and a diffractive index Δ_{n_λ}, is modeled as $\phi(x'', y'', \lambda) = e^{\frac{j2\pi\Delta_{n_\lambda} h(x'', y'')}{\lambda}}$.

Following the propagation and DOE effect operators single-doe propagation of a wavefront $u_0(\bar{x}, \bar{y}, \lambda)$ to a distance z_1 is modeled as:

$$u_1(x'', y'', \lambda) = \mathcal{P}\{u_0(\bar{x}, \bar{y}, \lambda), z_1\}. \quad (3)$$

where $u_1(x'', y'', \lambda)$ is the wavefront before the coding effect of the DOE. Then, propagating the encoded wavefront with the effect of the DOE to the sensor, a new wavefront is obtained:

$$u_2(x', y', \lambda) = \mathcal{P}\{\phi_1(x'', y'', \lambda) u_1(x'', y'', \lambda), z_2\}. \quad (4)$$

On the other hand, to describe a double-DOE architecture, as shown in Fig. 1, the single-DOE model is employed by adding a second DOE and another propagation to the sensor. The double phase encoded wavefront propagated to the sensor is modeled as:

$$u_3(x, y, \lambda) = \mathcal{P}\{\phi_2(x', y', \lambda) u_2(x', y', \lambda), z_3\} \quad (5)$$

The advantage of double-phase coding is summarized in the following lemma.

Lemma 1 *Let $\hat{u}_0(\bar{x}, \bar{y}, \lambda) = u_0(\bar{x} - a, \bar{y} - b, \lambda)$ be the source incident wavefront and the respective output of the double phase coding system $\hat{u}_3(x, y, \lambda)$, and $u_3(x, y, \lambda)$ the output for $u_0(\bar{x}, \bar{y}, \lambda)$. Then, given that the $\hat{u}_3(x, y, \lambda) \neq u_3(x-a, y-b, \lambda)$, it is possible to conclude that the double-phase codification system is shift-variant.*

Proof: The proof is contained in the supplementary material.

In this work, the sensor effect is considered for two cases; the first one is when the sensor pitch is equal to the desired image pitch. The second one is when the pitch size is bigger yielding low-resolution measurements. Remarking that in both cases is employed an RGB sensor to capture the spectral information.

2.1. Sensing process of spectral imaging

Considering $u_3(x, y, \lambda)$ the wavefront that arrives at a high-resolution spatial sensor with $M \times N$ pixels with a pitch of Δ, the discrete measurement in RGB can be defined as

$$\mathbf{G}(m, n, c) \propto \int_{(n)\Delta}^{(n+1)\Delta} \int_{(m)\Delta}^{(m+1)} \int_{\lambda} \kappa_c(\lambda)|u_3(x, y, \lambda)|d\lambda dxdy, \tag{6}$$

where $n \in \{0, ..., N-1\}$, $m \in \{0, ..., M-1\}$ and $\kappa_c(\lambda)$ is the wavelength sensitivity per channel c of the sensor.

2.2. Sensing process of low-spatial resolution spectral imaging

Considering a low-resolution spatial sensor with $L \times K$ pixels, where $L < M$ and $K < N$, and a low-resolution pitch of $\hat{\Delta}$, such that $\hat{\Delta} >> \Delta$, the discrete measurement in RGB can be defined as

$$\mathbf{G}_{LR}(k, l, c) \propto \int_{(l)\hat{\Delta}}^{(l+1)\hat{\Delta}} \int_{(k)\hat{\Delta}}^{(k+1)\hat{\Delta}} \int_{\lambda} \kappa_c(\lambda)\, |u_3(x, y, \lambda)|\, d\lambda dxdy, \tag{7}$$

where $l \in \{0, ..., L-1\}$, and $k \in \{0, ..., K-1\}$ and $\kappa_c(\lambda)$ is the wavelength sensitivity per channel c of the sensor.

3. Data Driven Optical Design

We propose to optimize the height maps of the two DOEs along with the reconstruction algorithm as shown in Fig. 2. First, we parametrize both height maps $\mathbf{H}_1$ and $\mathbf{H}_2$ via the Zernike polynomials basis. This is suitable parametrization since every element of the basis models a given optical aberration. Here, the coefficients of each term are the trainable parameter. Consequently, employing the Noll sequence [20], $\mathbf{H}_i$, for $i = 1, 2$ is defined as

$$\mathbf{H}_i = \sum_{k=1}^{\eta} \beta_i^{(k)} \mathbf{Z}^{(k)} \tag{8}$$

where η refers to the number of employed Zernike polynomials, $\boldsymbol{\beta}_i = [\beta_i^{(1)}, \ldots, \beta_i^{(\eta)}]$ are the coefficients of each polynomial for the i-th DOE and $\mathbf{Z}^{(k)}$ is the k-th Zernike polynomials.

Then, consider the discrete vectorized spectral image $\mathbf{u}^{MNL}$ where M, N are the spatial dimensions and the L are the number of spectral bands, the double DOE optical sensing is modeled with the operator $\mathcal{A}_{\mathcal{B}} : \mathbb{R}^{MNL} \rightarrow \mathbb{R}^{MN3}$ such that $\mathbf{g} = \mathcal{A}_{\mathcal{B}}(\mathbf{u})$, where $\mathcal{B} = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2\}$ and $\mathbf{g} \in \mathbb{R}^{MN3}$ are the coded measurements. Then, we aim to optimize jointly $\mathcal{B}$ with a decoding network that solves the inverse problem $\mathcal{D}_\theta$ with trainable parameters θ. Consider the spectral image dataset $\mathcal{U} = \{\mathbf{u}_t\}_{t=1}^{T}$, where T is the number of training examples. We formulate the E2E optimization as

$$\{\hat{\mathcal{B}}, \hat{\theta}\} = \arg\min_{\mathcal{B},\theta} \frac{1}{T} \sum_{t=1}^{T} \mathcal{L}_{rec}(\mathcal{D}_\theta(\mathcal{A}_{\mathcal{B}}(\mathbf{u}_t)), \mathbf{u}_t), \tag{9}$$

where $\mathcal{L}_{rec}$ is the recovery loss function.

3.1. Recovery Unrolling Network

To recover the spectral image from the coded measurements, an unrolling network is employed. This approach is inspired by traditional convex optimization formulations, which aim to solve a regularized data fidelity cost function. Consider the recovery of the t-th spectral image $\mathbf{u}_t$ from the measurements $\mathbf{g}_t$. The optimization problem to solve this inverse problem is

$$\hat{\mathbf{u}}_t = \arg\min_{\mathbf{u}_t} E(\mathbf{u}_t) + \tau R(\mathbf{u}_t) \tag{10}$$

where $E(\mathbf{u}_t)$ denotes the data fidelity cost function, τ is a regularization hyperparameter and $R(\mathbf{u}_t)$ is the regularization function that constrains the search space of the solution into a feasible set based on prior knowledge of the spectral image. A common choice for the data fidelity term is the ℓ_2 norm thus, we have that $E(\mathbf{u}_t) = \frac{1}{2}\|\mathbf{g}_t - \mathcal{A}_t(\mathbf{u}_t)\|_2^2$. We can re-formulate the optimization problem in (10) by including an auxiliary variable $\mathbf{v}_t$ following a half-quadratic splitting [11]. This method aims to split the optimization problem into two subproblems, one related to the data fidelity term and the other to the prior term. Thus, the (10) is re-written as

$$\begin{aligned} \{\hat{\mathbf{u}}_t, \hat{\mathbf{v}}_t\} = \arg\min_{\mathbf{u}_t, \mathbf{v}_t} \frac{1}{2}\|\mathbf{g}_t - \mathcal{A}_t(\mathbf{u}_t)\|_2^2 + \tau R(\mathbf{v}_t) \\ \text{subject to } \mathbf{v}_t = \mathbf{u}_t \end{aligned} \tag{11}$$

this constrained optimization problem is cast as a regularized problem as

$$\begin{aligned} \{\hat{\mathbf{u}}_t, \hat{\mathbf{v}}_t\} = \arg\min_{\mathbf{u}_t, \mathbf{v}_t} \frac{1}{2}\|\mathbf{g}_t - \mathcal{A}_t(\mathbf{u}_t)\|_2^2 + \tau R(\mathbf{v}_t) + \\ \mu\|\mathbf{v}_t - \mathbf{u}_t\|_2^2, \end{aligned} \tag{12}$$

where μ is another regularization parameter. Thus, (12) is solved in an alternating manner between $\mathbf{u}_t$ and $\mathbf{v}_t$ iteratively from $i = 1, \ldots, I$ where I is the number of iterations

$$\hat{\mathbf{v}}_t^{(i+1)} = \arg\min_{\mathbf{v}_t} \tau R(\mathbf{v}_t) + \mu\|\mathbf{v}_t - \mathbf{u}_t^{(i)}\|_2^2, \tag{13}$$

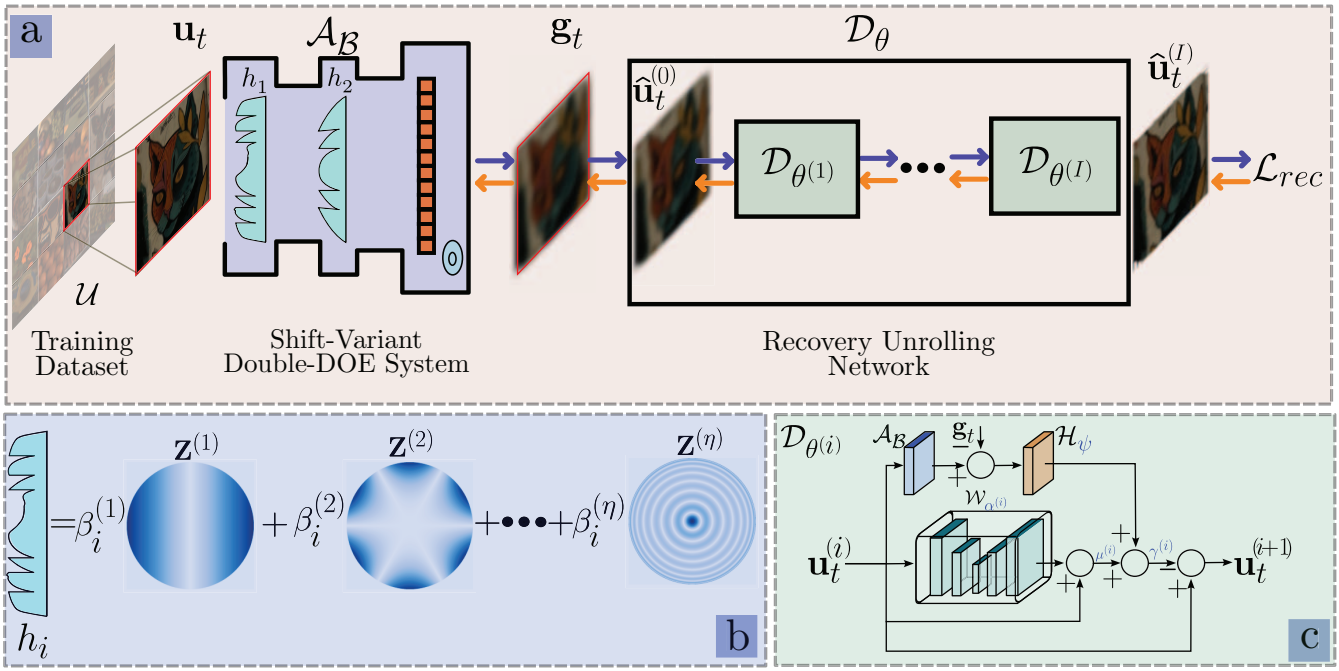

Figure 2. (a) General pipeline of the proposed End-to-End optimization of both heightmaps as coding step and the unrolling algorithm as decoding step. (b) Parametrization of the height maps as a linear combination of η Zernike polynomials. (c) Unrolling network stage where in blue are denoted the trainable parameters.

$$\hat{\mathbf{u}}_t^{(i+1)} = \arg\min_{\mathbf{u}_t} \frac{1}{2}\|\mathbf{g}_t - \mathcal{A}_t(\mathbf{u}_t)\|_2^2 + \mu\|\hat{\mathbf{v}}_t^{(i+1)} - \mathbf{u}_t\|_2^2. \tag{14}$$

The problem (13) is solved via a proximal step into the function $R(\cdot)$. Traditionally, this regularization function is hand-crafted and selected by the expertise of the user, however, this function fails to model the whole structure of the underlying signal in practical scenarios. Therefore, the unrolling algorithms aim to solve (13) via a deep neural network that learns priors from the training dataset denoted by $\mathcal{W}_{\alpha^{(i)}}$. Therefore, update of the of the variable $\mathbf{v}_t$ is

$$\hat{\mathbf{v}}_t^{(i+1)} = \mathcal{W}_{\alpha^i}(\hat{\mathbf{u}}_t^{(i)}), \tag{15}$$

where α^i are the trainable parameters. Then, the $\mathbf{u}_t$ subproblem (14) is solved via a gradient descent approach, yielding

$$\hat{\mathbf{u}}_t^{(i+1)} = \hat{\mathbf{u}}_t^{(i)} - \gamma^{(i)}\Big(-\mathcal{A}_{\mathcal{B}}^*(\mathbf{g}_t - \mathcal{A}_{\mathcal{B}}(\hat{\mathbf{u}}_t^{(i)})) - \mu_t(\hat{\mathbf{v}}_t^{(i+1)} - \mathbf{u}^{(i)})\Big), \tag{16}$$

where $\mathcal{A}_{\mathcal{B}}^*$ is the adjoint operator of the sensing process and $\gamma^{(i)}$ is the gradient step size at the ith iteration.. To model the adjoint sensing operator, a deep neural network that computes the adjoint mapping is used. This network is denoted as $\mathcal{H}_\psi$ where ψ are the trainable weights. Replacing the expression for $\mathbf{v}_t^{(i+1)}$, we obtain

$$\hat{\mathbf{u}}_t^{(i+1)} = \hat{\mathbf{u}}_t^{(i)} - \gamma^{(i)}\Big(-\mathcal{H}_\psi(\mathbf{g}_t - \mathcal{A}_{\mathcal{B}}(\hat{\mathbf{u}}_t^{(i)})) - \mu^{(i)}(\mathcal{W}_{\alpha^{(i)}}(\hat{\mathbf{u}}_t^{(i)}) - \hat{\mathbf{u}}_t^{i})\Big). \tag{17}$$

This update rule is computed I times. Additionally, the parameters $\gamma^{(i)}$ and $\mu^{(i)}$ which are usually selected via trial and error, with the unrolling formulation, these parameters are also trained. We can express every stage of the unrolling network as $\hat{\mathbf{u}}_t^{(i+1)} = \mathcal{D}_{\theta^{(i)}}(\hat{\mathbf{u}}_t^{(i)})$ where

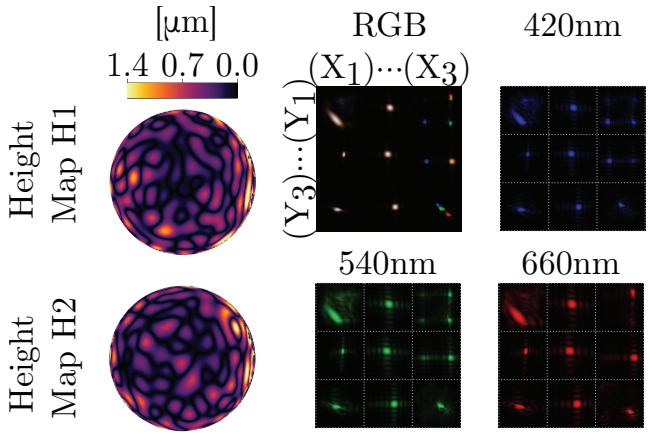

Figure 3. Learned Height-maps and point spread functions of the system in various spatial points of the sensor for spectral imaging task

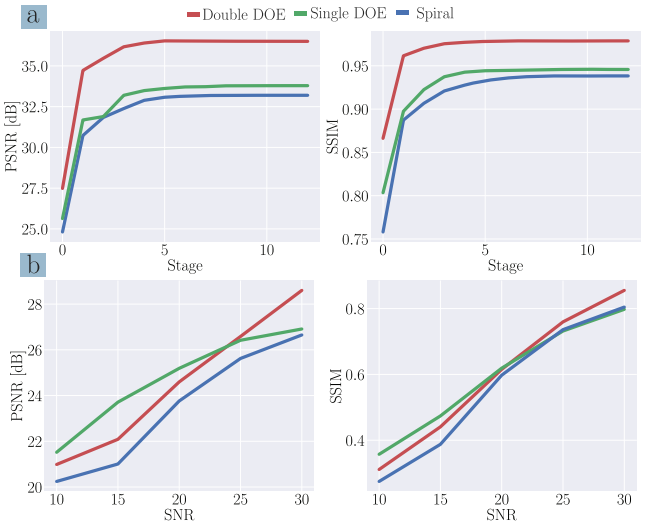

Figure 4. Numerical results over the test dataset for spectral imaging task

$\theta^{(i)} = \{\gamma^{(i)}, \alpha^{(i)}, \mu^{(i)}\}$ are all the trainable parameters at each stage. For the initialization of the unrolling algorithm we employ $\mathbf{u}_t^{(0)} = \mathcal{H}_\psi(\mathbf{g}_t)$

4. Simulations Experiments

To train the end-to-end network, the ARAD dataset was employed [8], where 900 images were used for training and 50 for testing with 512×512 pixel and 31 spectral bands. The framework was programmed in Tensorflow/Keras [1] API. For the unrolling network, the proximal step subnetwork $\mathcal{W}_{\alpha^{(i)}}$ chosen was the network proposed in [9] which is based on a linear mixture model of spectral images. The optimization problem in (9) was solved using Adam optimizer [19] with its hyperparameters set as default for 100 epochs. The learning rate was halved every 50 epochs. To assess performance, two metrics were used: the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) [15]. As loss function $\mathcal{L}_{rec}$, the mean squared error was used, additionally, the loss function was computed at the end of each unrolling stage to improve the convergence of the network as suggested in [17]. The number of Zernike polynomials employed to train the DOEs height maps was set to $\eta = 200$. All the simulations were performed using a Nvidia GPU 3090 with 24 GB memory.

4.1. Application on Spectral Image Reconstruction

For this experiment, the spectral images were cropped to 128x128 pixels with 25 spectral bands for the spectral imaging task. The number of stages employed in the unrolling network was set to $I = 12$, some experiments were performed changing the number of stages parameters but after $I = 12$ the results did not improve.

The learned height maps for each of the DOE of the double-DOE architecture are presented in Fig.3, also the point spread functions, where X_n and Y_n with $n \in [1, 2, 3]$ represent different regions of the sensor, are presented, showing that the system varies in both space and spectrum.

In Fig. 4 can be seen that the proposed double-doe architecture beats the state-of-the-art Spiral DOE, and a single learnable DOE from the initialization stage to the final stage, having a gap in performance of up to 3 [dB]. An analysis of noisy scenarios was made, using signal-to-noise-ratio (SNR) of 10, 15, 20, 25, and 30 [dB] without retraining the network weights, the results of the evaluation are shown in Fig.4 where can be seen that even in low SNR scenarios, the proposal shows comparable results, and in scenarios with high SNR the proposed architecture gain in up to 2 [dB].

The visual results presented in Fig. 5 are a comparison of two images of the test dataset, the architectures are compared of PSNR and the Spectral Angle Mapper (SAM) of specific spatial points. Is important to note that a bigger number in PSNR leads to better reconstruction performance and, a smaller SAM implies better spectral reconstruction of signatures. These visual results also show that the Double DOE architecture leads to reconstruction with fewer artifacts and better overall recovery.

4.2. Application on Spatial Super-resolution

In this subsection, the spectral images were kept to 512×512 and 9 equispaced spectral bands. The downsampled factor was set to 2, i.e. $\frac{M}{L} = 2$. Due to the high dimensionality of the dataset for this experiment, the number of stages of the unrolling network was set to $I = 3$. Learned height maps for each of the DOE of the double-DOE architecture are presented in Fig.6, also the point spread functions on various points of the sensor are presented, showing that the system varies spatially.

In Fig.7 the results in terms of PSNR and SSIM are shown for both noiseless and noisy scenarios. For the case of noiseless scenarios, the proposed double-doe architecture beats the state-of-the-art Spiral DOE, and a single learnable DOE in all stages, having a gap in performance of up to 3.5 [dB]. In the case of noisy scenarios, the analysis was conducted using SNR of 10, 15, 20, 25, and 30 [dB], and from Fig.7 b, can be concluded that the gap between double-DOE and single-DOE architectures in terms of PSNR is reduced while the SNR increases.

In the visual results presented in Fig. 8 the architectures are compared using PSNR and SAM of specific spatial points in two images of the test dataset. These visual results also show that the Double DOE architecture leads to reconstruction with better spatial definition and spectral recovery.

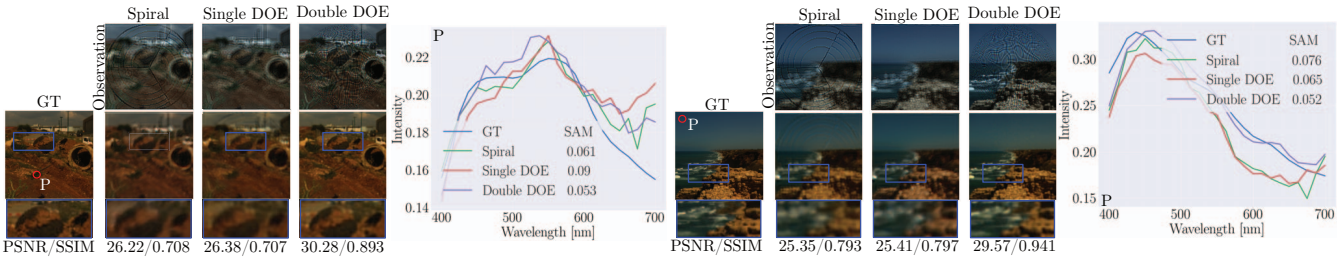

Figure 5. Visual results of two images from the test dataset, comparing measurements, reconstruction, spectral signature, and metrics for spectral reconstruction.

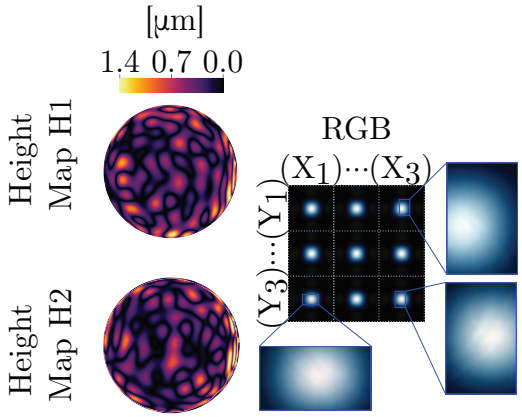

Figure 6. Learned Height-maps and point spread functions in RGB representation of the system in various points of the sensor for super-resolution task

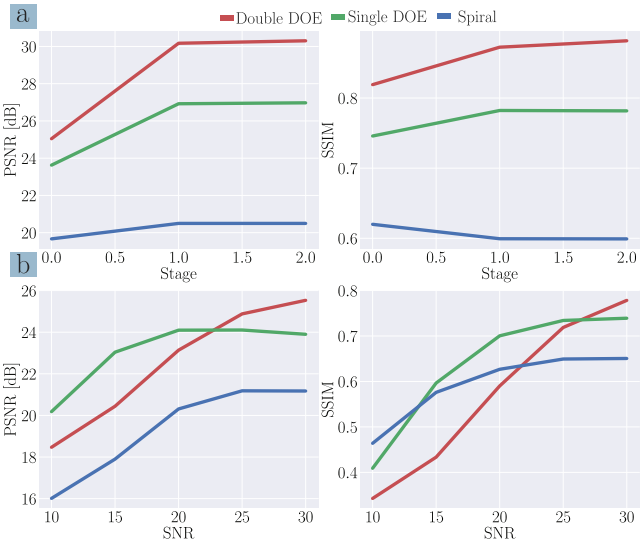

Figure 7. Numerical results over the test dataset for spatial super-resolution application

5. Conclusion

A shift-variant double-DoE architecture, suitable for multiple tasks in spectral images, such as recovery, and super-resolution has been proposed. This architecture allows high-fidelity spectral image reconstruction and it is feasible for other high-level tasks, leading to the possibility of spatial-spectral super-resolution images. Simulation results show that the proposed architecture can achieve better performance in spatial and spectral reconstruction compared with the shift-invariant DOE-based architectures. Specifically, the proposed architecture improves the reconstruction quality by up to 3 [dB].

6. Acknowledgement

This work was supported by ICETEX and MINCIENCIAS through the CTO 2022-0716 Sistema ópticocomputacional tipo pushbroom en el rango visible e infrarrojo cercano (VNIR), para la clasificación de frutos cítricos sobre bandas transportadoras mediante aprendizaje profundo, desarrollado en alianza con citricultores de Santander, under Grant 8284

References

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pages 265–283, 2016.

[2] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115, 2014.

[3] Henry Arguello, Jorge Bacca, Hasindu Kariyawasam, Edwin Vargas, Miguel Marquez, Ramith Hettiarachchi, Hans Garcia, Kithmini Herath, Udith Haputhanthri, Balpreet Singh Ahluwalia, et al. Deep optical coding design in computational imaging: a data-driven framework. *IEEE Signal Processing Magazine*, 40(2):75–88, 2023.

[4] Henry Arguello, Samuel Pinilla, Yifan Peng, Hayato Ikoma, Jorge Bacca, and Gordon Wetzstein. Shift-variant color-coded diffractive spectral imaging system. *Optica*, 8(11):1424–1434, 2021.

[5] Jorge Bacca, Emmanuel Martinez, and Henry Arguello. Computational spectral imaging: a contemporary overview. *JOSA A*, 40(4):C115–C125, 2023.

[6] Seung-Hwan Baek, Hayato Ikoma, Daniel S Jeon, Yuqi Li, Wolfgang Heidrich, Gordon Wetzstein, and Min H

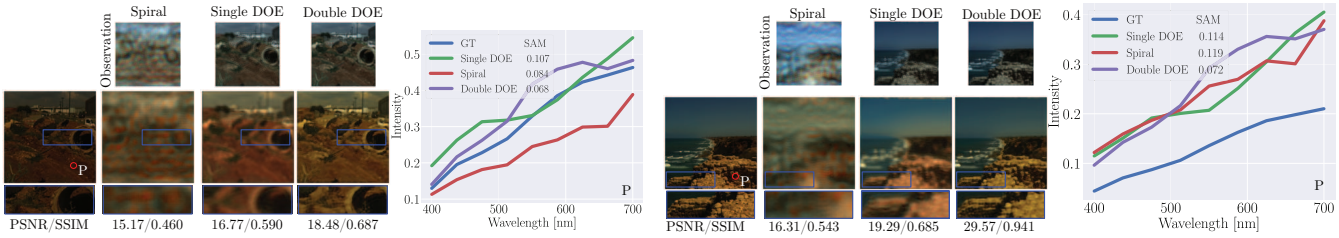

Figure 8. Visual results of two images from the test dataset, comparing measurements, reconstruction, spectral signature, and metrics for spatial super-resolution reconstruction.

Kim. Single-shot hyperspectral-depth imaging with learned diffractive optics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2651–2660, 2021.

[7] Simon Denman, Todd Lamb, Clinton Fookes, Vinod Chandran, and Sridha Sridharan. Multi-spectral fusion for surveillance systems. *Computers & Electrical Engineering*, 36(4):643–663, 2010.

[8] Arad et al. Ntire 2022 spectral recovery challenge and data set. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 862–880, 2022.

[9] Tatiana Gelvez, Jorge Bacca, and Henry Arguello. Interpretable deep image prior method inspired in linear mixture model for compressed spectral image recovery. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 1934–1938. IEEE, 2021.

[10] Tatiana Gelvez, Hoover Rueda, and Henry Arguello. Joint sparse and low rank recovery algorithm for compressive hyperspectral imaging. *Appl. Opt.*, 56(24):6785–6795, Aug 2017.

[11] Donald Geman and Chengda Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE transactions on Image Processing*, 4(7):932–946, 1995.

[12] Joseph W Goodman. Introduction to fourier optics. *Introduction to Fourier optics, 3rd ed., by JW Goodman. Englewood, CO: Roberts & Co. Publishers, 2005*, 1, 2005.

[13] Megandhren Govender, K Chetty, and Hartley Bulcock. A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water Sa*, 33(2), 2007.

[14] Nathan A. Hagen and Michael W. Kudenov. Review of snapshot spectral imaging technologies. *Optical Engineering*, 52(9):1 – 23, 2013.

[15] Alain Horé and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369, 2010.

[16] Tao Huang, Weisheng Dong, Xin Yuan, Jinjian Wu, and Guangming Shi. Deep gaussian scale mixture prior for spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16216–16225, 2021.

[17] Roman Jacome, Jorge Bacca, and Henry Arguello. D 2 uf: Deep coded aperture design and unrolling algorithm for compressive spectral image fusion. *IEEE Journal of Selected Topics in Signal Processing*, 2022.

[18] Daniel S Jeon, Seung-Hwan Baek, Shinyoung Yi, Qiang Fu, Xiong Dun, Wolfgang Heidrich, and Min H Kim. Compact snapshot hyperspectral imaging with diffracted rotation. *Association for Computing Machinery (ACM)*, 2019.

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[20] Vasudevan Lakshminarayanan and Andre Fleck. Zernike polynomials: a guide. *Journal of Modern Optics*, 58(7):545–561, 2011.

[21] Chao Li, Souleymane Balla-Arabé, and Fan Yang. Embedded multi-spectral image processing for real-time medical application. *Journal of Systems Architecture*, 64:26 – 36, 2016. Real-Time Signal Processing in Embedded Systems.

[22] Xing Lin, Gordon Wetzstein, Yebin Liu, and Qionghai Dai. Dual-coded compressive hyperspectral imaging. *Optics letters*, 39(7):2044–2047, 2014.

[23] Yang Liu, Xin Yuan, Jinli Suo, David J Brady, and Qionghai Dai. Rank minimization for snapshot compressive imaging. *IEEE transactions on pattern analysis and machine intelligence*, 41(12):2990–3006, 2018.

[24] Miguel Marquez, Pablo Meza, Henry Arguello, and Esteban Vera. Compressive spectral imaging via deformable mirror and colored-mosaic detector. *Optics express*, 27(13):17795–17808, 2019.

[25] Xin Miao, Xin Yuan, Yunchen Pu, and Vassilis Athitsos. l-net: Reconstruct hyperspectral images from a snapshot measurement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4059–4069, 2019.

[26] Xin Miao, Xin Yuan, and Paul Wilford. Deep learning for compressive spectral imaging. In *Digital Holography and Three-Dimensional Imaging*, pages M3B–3. Optical Society of America, 2019.

[27] Edwin Vargas, Henry Arguello, and Jean-Yves Tourneret. Spectral image fusion from compressive measurements using spectral unmixing. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 1–5. IEEE, 2017.

[28] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Appl. Opt.*, 47(10):B44–B51, Apr 2008.

[29] Chang Xu, Tingfa Xu, Ge Yan, Xu Ma, Yuhan Zhang, Xi Wang, Feng Zhao, and Gonzalo R Arce. Super-resolution compressive spectral imaging via two-tone adaptive coding. *Photonics Research*, 8(3):395–411, 2020.

[30] Xin Yuan, David J Brady, and Aggelos K Katsaggelos. Snapshot compressive imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 38(2):65–88, 2021.