SDI-PixCar: Single Snapshot Compressive Spectral Depth Imaging via Pixeled Carrier

Kevin Arias, Jhon Lopez, Miguel Marquez, Carlos Hinojosa, Henry Arguello
Universidad Industrial de Santander
Bucaramanga, Colombia
kevin.arias@correo.uis.edu.co

Abstract

Single snapshot spectral-depth imaging (SDI) recovery is an open problem in the computer vision community. Few works have been tackled the SDI problem by introducing systems that rely either on multiple sensors placed in tandem or low light-efficient diffractive optics approach. In this work, we propose a single snapshot compressive spectral depth imager via a pixeled carrier (SDI-PixCar). Specifically, SDI-PixCar relies on a dual-dispersive sensing and a pixelated carrier illumination approach to generate a compressed measurement that encodes spectral and depth features. The spectral and depth reconstruction is performed with a convolutional neural network (CNN) based on U-net. Extensive simulations demonstrate that the proposed SDI-PixCar represents an efficient alternative to estimate the SDI.

1. Introduction

Spectral imaging sensors collect spectral information at every spatial location of a scene. The acquired high dimensional data are commonly regarded as a three-dimensional structure also known as spectral image (SI) $\mathbf{X} \in \mathbb{R}^{N \times M \times L}$, where (x, y) are its spatial coordinates, and (λ) represents the spectral wavelengths [21]. The rich spectral information contained in the SI is an essential for several applications such as precision agriculture, monitoring and clustering [5, 14, 24]. There are different techniques for SI acquisition, such as whiskbroom or pushbroom scanning systems which have limitations of spatial-spectral resolution and acquisition times [6, 13]. As an alternative, instantaneous spectral imaging systems such as the coded-aperture spectral imaging system (CASSI) [2, 15, 16, 23] or dual-disperser coded-aperture spectral imaging system (DD-CASSI) [10, 17] can sense spatial-spectral information as 2D compressed projections, reducing the amount of data. On the other hand, 3D shape reconstruction techniques by projecting high power light onto scene are useful in some areas such as semantic segmentation, autonomous car driving and security [1, 12, 22]. In this way, depth information of a scene can be estimated by projecting a set of structured light (SL) patterns over the surface object. The most popular SL works are based on the phase-shifting structured amplitude theory that aims to disambiguate the depth information from a set of at least three SL projections [27]. To increase the spectral resolution without resort to scanning protocols, researchers have explored compressive imaging (CI) theory [2, 3] to propose novelty imagers and reconstruction paradigms with the capability of estimate the spectral and depth information from a reduced set of two-dimensional projections. This framework is commonly known as compressive structured light spectral depth imaging (CSL-SDI), and their sensing paradigm aims to modulate and compress, at the same time the spectral and depth information. Conventionally, the CLS-SDIs carry out this modulation/compression process by taking advantage of optical elements such as light projectors (LP) [19], binary coded apertures (CA), color-coded apertures (CCA) [4], CFMD, and dispersive elements [8]. For the spectral sensing, the most popular compressive sensing geometry is CASSI. Although the CSL-SDI systems enable the acquisition of the SDI with a reduced set of measurements, these systems are yet limited by the minimum total of projections requested for the PSA approach. This paper proposes a CSL-SDI system that acquires compressed measurements that preserves the spectral and 3D-shape features by multiplexing a grayscale and dispersed-coded-dispersed projection of the scene into a single sensor. The dual-dispersive sensing geometry and the pixelated carrier interferogram sensing theory are employed to propose a CSL-SDI system that captures a dispersed-coded-dispersed measurement resulting from a monocular system employing a single sensor and prism. The optical architecture does not rely on a camera array but a single imaging sensor with a single coded aperture and a double Amici prism as a dispersive element. Moreover, the structured light patterns can be designed in a way that the whole set of uncorrelated compressed measurements can be used to reconstruct the spectral information. Further, we use a two-arm U-net-based neural network to recover the spectral datacube and the phase information [7].

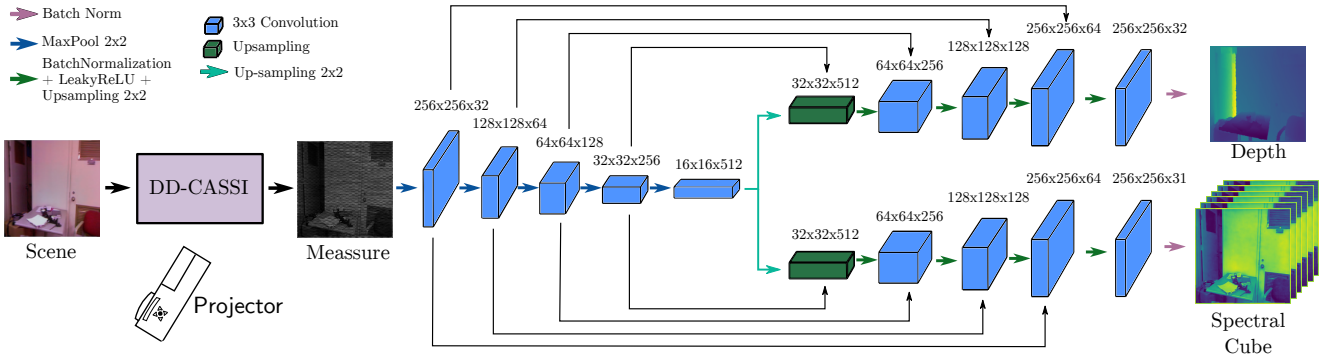


Figure 1. SDI-PixCar network. The first step we extract the same features from the compressed image and then the features are used separately in a U-net decoder to estimated the spectral and depth information.

2. Proposed method

We are interested in the spectrum and phase estimation from single shot. Our general idea is to project on an object a single pattern of structured light and acquire a set of projections that compresses spectral and depth information from DD-CASSI architecture. Thus, We designed a light projection geometry that consolidates four different patterns enabling depth recovery guarantees from a single sensing. Then, the DD-CASSI sensing model is used to modulate the spectrum of the patterned object in a single shot.

2.1. Optical Encoder

An object $f(x, y, \lambda)$ with spatial coordinates (x, y) is illuminated on its surface by a projector with structured light patterns generated from a pixelated carrier [26]. The result reflected light can be mathematically expressed as

$$f_p(x, y, \lambda) = a(x, y, \lambda) + f(x, y, \lambda)\,[b(x, y, \lambda) + \cos(\varphi(x, y) + \omega_0(x, y)), \tag{1}$$

where $\varphi(x, y)$ is the object phase information, $\omega_0(x, y)$ is the spatial carrier, $b(x, y, \lambda)$ is a balance function that avoids that $f_p(x, y, \lambda)$ achieve negative values and $a(x, y, \lambda)$ represents the background. The bi-dimensional pixelated carrier $\omega_0(x, y)$ gives for each location the phase step introduced by the pixelated phase mask which encoded phase/shape information into its deformation on object.

4-in-1 pattern design. Particularly, we project on an object a pixelated carrier and acquire the reflected light in a single-shot under the DD-CASSI sensing model avoiding the multi-shot acquisition of use a temporal carrier also known as linear carrier where the measurements are conformed by t shots. In detail, the pixeled carrier is built from multiple linear carriers to consolidate several different linear spatial patterns into one [20]. Specifically, we use four linear carriers to conform our pixelated carrier $\mathbf{W} \in \mathbb{R}^{2\times 2}$ with window structure $\mathbf{W} = \begin{bmatrix} 0 & \pi/2 \\ \pi & 3\pi/2 \end{bmatrix}$ that is repeated along of rows and columns of a matrix structure, such that

$$\bar{\mathbf{W}} = \mathbf{O} \otimes \mathbf{W}, \tag{2}$$

where $\bar{\mathbf{W}} \in \mathbb{R}^{N\times M}$ is the pixelated carrier, $\mathbf{O} \in \mathbb{R}^{N/2, N/2}$ is a ones matrix and $\otimes$ denotes the Kronecker product.

DD-CASSI architecture. Our acquisition process in Fig. 1 is modeled by the DD-CASSI architecture [10] that modulates the patterned object through a coded aperture under dual-dispersion where two dispersers are symmetrically placed on the two sides of the coded aperture to overcome the shear effect of the SD-CASSI [9]. The dual-dispersed modulated version of the object can be formulated as

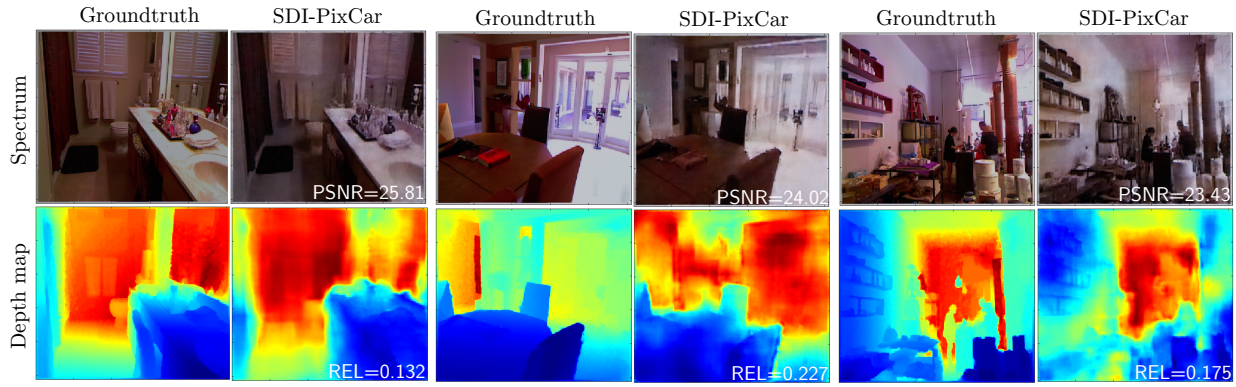
$$\tilde{g}(x', y', \lambda') = \iint h(x' - x'' - S(\lambda'), y' - y'', \lambda') \times f_p(x'', y'', \lambda') k(x'', y'') dx'' dy'', \tag{3}$$

$$\text{with }, k(x, y) = \frac{1}{2} \sum_{i'_1, i_2} \Pi\left(\frac{x - i'_1 \Delta_c}{\Delta_c}\right) \Pi\left(\frac{y - i_2 \Delta_c}{\Delta_c}\right),$$

where Π represents the rectangle function, S_λ is a wavelength-dependent function, $h(x - x' - S(\lambda), y - y', \lambda)$ accounts for the impulse response of the optical system, Δ_c denoting the pixel size, Δ_λ denoting the spectral bandwidth, $\mathbf{C} \in \mathbb{R}^{N\times M+L-1}$ is a binary-coded aperture with $C_{i'_1, i_2} \in \{0, 1\}$, and $i'_1 = \{0, \ldots, (N_x + N_\lambda - 2)\}$ and $i_2 = \{0, \ldots, N_y - 1\}$ indexing the rows and columns, respectively. Then, the dispersed-modulated wavefront is propagated to the prism to correct the dispersion effect. Finally, the dual-dispersed modulated wavefront is propagated and focused into a detector array to obtain the following measurements

$$g(x, y) = \iiint h(x - x' + S(\lambda), y - y', \lambda) \times \tilde{g}(x', y', \lambda') dx' dy' d\lambda, \tag{4}$$

where $g(x, y)$ denotes the spectral + depth compressed measurements. Mathematically, the linear discrete sensing from DD-CASSI system according to Eq.4 is formulated as



(a) Scene 1 (b) Scene 2 (c) Scene 3

Figure 2. Spectral and depth estimation results of the SDI Pix-Car network for three different scenes. The average metric of the full data using the SDI-PixCar network are SSIM= 0.813 and PSNR=23.81$[dB]$ for spectral estimation and SSIM = 0.892, REL = 0.186 and MSE = 0.014 for depth estimation.

$$G_{i_1,i_2} = \sum_{i_3=0}^{L-1} C_{i_1,(i_2-i_3)} F_{p\ i_1,i_2,i_3} + \epsilon_{i_1,i_2}, \tag{5}$$

where $\mathbf{G} \in \mathbb{R}^{N\times M}$ denotes the compressed measurements, $\mathbf{F}_p \in \mathbb{R}^{M\times N\times L}$ is the 3D tensor that contains the spectral response of the scene for the single structured light projection and $\epsilon \in \mathbb{R}^{N\times M}$ represents the additive noise. The linear sensing model can be expressed as $\mathbf{g} = \mathbf{H}\mathbf{f}_p + \boldsymbol{\epsilon}$, where $\mathbf{g} \in \mathbb{R}^m$ is compressed measurement in vector form, $\mathbf{H} \in \mathbb{R}^{m\times n}$ is the sensing matrix, and $\boldsymbol{\epsilon} \in \mathbb{R}^m$ is the noise.

2.2. Spectrum + depth deep network

We use the U-NET network architecture with a dual decoder design. The features are shared for both spectrum and depth but are independently decoded by two arms, one for depth and other for the spectrum as shown in Fig. 1. Thus, the spectral images $\mathbf{f}$ and depth map $\boldsymbol{\varphi}$ are simultaneously estimated from measurements $\mathbf{g}$ as follows

$$\hat{\mathbf{f}}, \hat{\boldsymbol{\varphi}} = \mathcal{N}_{SDI-PixCar}(\mathbf{g}) \tag{6}$$

Loss function. the network learning is addressed from a joint loss, both the structural similarity index measure (SSIM) [11] and Mean absolute error (MAE) are employed to recover the spectrum and phase information

$$\begin{aligned}\mathcal{L} = \lambda_s &\left[\left(1 - SSIM\left(\mathbf{f}, \hat{\mathbf{f}}\right)\right) + MAE\left(\mathbf{f}, \hat{\mathbf{f}}\right)\right] \\ &+ \left[\left(1 - SSIM\left(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}\right)\right) + MAE\left(\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}\right)\right]\end{aligned} \tag{7}$$

Training stage. We optimize the spectral + depth network under a preset random coded aperture via backpropagation by solving for the following minimization problem

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} \mathcal{L}\left(\left\{\hat{\varphi}\left(\boldsymbol{\theta}\right), \hat{f}\left(\boldsymbol{\theta}\right)\right\}, \{\varphi, f\}\right), \tag{8}$$

where $\boldsymbol{\theta}$ is the set of network parameters. Note that, the spectrum and depth estimation share feature encoding reducing the number of trainable parameters $\boldsymbol{\theta}$ of the neural network. Additionally, the shared use of the encoder can be seen as a helpful regularization for the weights.

3. Simulation results

We use a synthetic spectral and depth dataset generated from the public NYU RGB-D dataset [18]. This dataset is composed of 1499 files, where each one contains an spectral secene and the depth map information of the scene. In the training, we resized datacubes to $N \times M = 256 \times 256$ and $L = 31$. Then, the dataset is split into 869 images for training and 630 for testing. The hyperparameters were set up as follow: batch number 16, epochs number 100, Adam optimizer with learning rate of $1e^{-4}$ and regularization parameter $\lambda_s = 1$. Figure 2 shows the spectral and depth reconstruction results obtained for three representative spectral-depth images. These results prove that SDI-PixCar is capable to reconstruct the spectral and depth information from a single compressed measurement. Specifically, the SDI-PixCar achieves an average Peak signal-to-noise ratio (PSNR) and SSIM of 23.81 dB and 0.813, respectively, for the spectral datacube reconstruction. For phase information, the SDI-PixCar achieves an average SSIM, REL [25] and MSE of 0.892, 0.186 and 0.014, respectively.

4. Conclusions

We have developed the SDI-PixCar that enables spectral-depth image retrieval from a single compressed measurement. Numerical simulations verify the feasibility of the proposed paradigm with a dual-dispersive spectral system and a pixeled carrier structured light pattern. The SDI-PixCar approach is highly potential for SDI applications in ultrafast imaging, i.e., in scenarios where acquiring multiple shots per frame is not feasible.

References

[1] Mahbubul Alam, Manar D Samad, Lasitha Vidyaratne, Alexander Glandon, and Khan M Iftekharuddin. Survey on deep neural networks in speech and vision systems. *Neurocomputing*, 417:302–321, 2020. 1

[2] Gonzalo R Arce, David J Brady, Lawrence Carin, Henry Arguello, and David S Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115, 2013. 1

[3] Henry Arguello and Gonzalo R Arce. Code aperture optimization for spectrally agile compressive imaging. *JOSA A*, 28(11):2400–2413, 2011. 1

[4] Henry Arguello and Gonzalo R Arce. Colored coded aperture design by concentration of measure in compressive spectral imaging. *IEEE Transactions on Image Processing*, 23(4):1896–1908, 2014. 1

[5] Kevin Arias, Edwin Vargas, and Henry Arguello. Hyperspectral and multispectral image fusion based on a non-locally centralized sparse model and adaptive spatial-spectral dictionaries. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pages 1–5. IEEE, 2019. 1

[6] Kevin Arias, Edwin Vargas, Fernando Rojas, and Henry Arguello. Fusion of hyperspectral and multispectral images based on a non-locally centralized sparse model of abundance maps. *Tecnura*, 24(66):62–75, 2020. 1

[7] Seung-Hwan Baek, Hayato Ikoma, Daniel S Jeon, Yuqi Li, Wolfgang Heidrich, Gordon Wetzstein, and Min H 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. 1

[8] Seung-Hwan Baek, Incheol Kim, Diego Gutierrez, and Min H Kim. Compact single-shot hyperspectral imaging using a prism. *ACM Transactions on Graphics (TOG)*, 36(6):1–12, 2017. 1

[9] Xun Cao, Tao Yue, Xing Lin, Stephen Lin, Xin Yuan, Qionghai Dai, Lawrence Carin, and David J Brady. Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, 33(5):95–108, 2016. 2

[10] Michael E Gehm, Renu John, David J Brady, Rebecca M Willett, and Timothy J Schulz. Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics express*, 15(21):14013–14027, 2007. 1, 2

[11] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE, 2010. 3

[12] Wenyi Huang, Junsheng Cheng, Yu Yang, and Gaoyuan Guo. An improved deep convolutional neural network with multiscale information for bearing fault diagnosis. *Neurocomputing*, 359:77–92, 2019. 1

[13] Qingli Li, Xiaofu He, Yiting Wang, Hongying Liu, Dongrong Xu, and Fangmin Guo. Review of spectral imaging technology in biomedical engineering: achievements and challenges. *Journal of biomedical optics*, 18(10):100901, 2013. 1

[14] Jhon Lopez, Carlos A Hinojosa, and Henry Arguello. Efficient subspace clustering of hyperspectral images using similarity-constrained sampling. *Journal of Applied Remote Sensing*, 15(3):036507, 2021. 1

[15] Miguel Marquez, Hoover Rueda-Chacon, and Henry Arguello. Compressive spectral light field image reconstruction via online tensor representation. *IEEE Transactions on Image Processing*, 29:3558–3568, 2020. 1

[16] Miguel Marquez, Hoover Rueda-Chacon, and Henry Arguello. Compressive spectral imaging via virtual side information. *IEEE Transactions on Computational Imaging*, 7:114–123, 2021. 1

[17] Jonathan Monsalve, Miguel Marquez, Iñaki Esnaola, and Henry Arguello. Compressive covariance matrix estimation from a dual-dispersive coded aperture spectral imager. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 2823–2827. IEEE, 2021. 1

[18] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012. 3

[19] Halina Rubinsztein-Dunlop, Andrew Forbes, Michael V Berry, Mark R Dennis, David L Andrews, Masud Mansuripur, Cornelia Denz, Christina Alpmann, Peter Banzer, Thomas Bauer, et al. Roadmap on structured light. *Journal of Optics*, 19(1):013001, 2016. 1

[20] Manuel Servin, J Antonio Quiroga, and Moises Padilla. *Fringe pattern analysis for optical metrology: theory, algorithms, and applications*. John Wiley & Sons, 2014. 2

[21] Gary A Shaw and Hsiao-Hua K Burke. Spectral imaging for remote sensing. *Lincoln Laboratory Journal*, 14(1):3–28, 2003. 1

[22] Guanzhong Tian, Liang Liu, JongHyok Ri, Yong Liu, and Yiran Sun. Objectfusion: An object detection and segmentation framework with rgb-d slam and convolutional neural networks. *Neurocomputing*, 345:3–14, 2019. 1

[23] Ashwin A Wagadarikar, Nikos P Pitsianis, Xiaobai Sun, and David J Brady. Spectral image estimation for coded aperture snapshot spectral imagers. In *Image Reconstruction from Incomplete Data V*, volume 7076, pages 9–23. Spie, 2008. 1

[24] Eric M Wood, Anna M Pidgeon, Volker C Radeloff, and Nicholas S Keuler. Image texture as a remotely sensed measure of vegetation structure. *Remote Sensing of Environment*, 121:516–526, 2012. 1

[25] Yicheng Wu, Vivek Boominathan, Huaijin Chen, Aswin Sankaranarayanan, and Ashok Veeraraghavan. Phasecam3d—learning phase masks for passive single view depth estimation. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2019. 3

[26] Peizheng Yan, Xiangwei Liu, Shuangle Wu, Fangyuan Sun, Qihan Zhao, and Yonghong Wang. Pixelated carrier phase-shifting shearography using spatiotemporal low-pass filtering algorithm. *Sensors*, 19(23):5185, 2019. 2

[27] Wei Yin, Qian Chen, Shijie Feng, Tianyang Tao, Lei Huang, Maciej Trusiak, Anand Asundi, and Chao Zuo. Temporal phase unwrapping using deep learning. *Scientific reports*, 9(1):1–12, 2019. 1