Demultiplexing Visible and Near-Infrared Information in Single-Sensor Multispectral Imaging

***Zahra Sadeghipoor*[1]*, Jean-Baptiste Thomas*[2]*, and Sabine Süsstrunk*[1]**
[1] ***École Polytechnique Fédérale de Lausanne (EPFL), Switzerland and*** [2] ***Université de Bourgogne, France***

Abstract

In this paper, we study a single-sensor imaging system that uses a multispectral filter array to spectrally sample the scene. Our system captures information in both visible and near-infrared bands of the electromagnetic spectrum. Due to manufacturing limitations, the visible filters in this system also transmit the NIR radiation. Similarly, visible light is transmitted by the NIR filter, leading to inaccurate mixed spectral measurements. We present an algorithm that resolves this issue by separating NIR and visible information. Our method achieves this goal by exploiting the correlation of multispectral images in both spatial and spectral domains. Simulation results show that the mean square error of the data corrected by our method is less than 1/20 of the error in sensor spectral measurements.

Introduction

Multispectral images are extensively used in several fields such as remote sensing, agriculture, product quality assessment, security and surveillance. The conventional multispectral imaging systems either use a filter wheel or a beam splitter with several detectors to capture different spectral channels of the scene [9]. These systems are costly and can take up to a few minutes to scan the scene. To overcome these difficulties, single-sensor multispectral acquisition systems that use spectral filter arrays (SFA) have been proposed [12].

In this paper, we study a single-sensor imaging system that captures eight spectral channels in visible and near-infrared (NIR) bands of the electromagnetic spectrum. The single-sensor acquisition is enabled by the fact that silicon sensors are inherently responsive in the wavelength range of 400-1100 nm (visible band: 400-750 nm; NIR: 750-1100 nm). Figure 1 illustrates the quantum efficiency curves for two silicon-based sensors. Single-sensor acquisition of visible and NIR information is studied by both the computational photography and the multispectral imaging communities [4, 6, 7, 8, 9, 10, 14].

One of the main issues with such a camera is that the filters in the SFA cannot be manufactured to have a precise spectral transmittance with no leakage. Visible filters usually transmit NIR radiation and NIR filters pass the visible light as well. This "spectral cross-talk" diminishes the value of capturing both visible and NIR bands by increasing the redundancy in the measurements. This deficiency in filter transmittance is imposed by current manufacturing limitations. For instance, the Fabry-Perot interferometer may show a harmonic, i.e. a second transmittance peak, further away from the main transmittance peak of the filter.

In this paper, we develop a signal-separation technique that estimates visible and NIR intensities from mixed sensor measurements. Our algorithm exploits the spatial and spectral correlations of data by computing the sparse representations of different channels. Sparse decomposition methods are used both in the source separation literature [2, 18] and in color and multispectral imaging [3, 5, 16]. Golbabaee *et al.* use sparse decomposition to reconstruct multispectral data from a few noisy measurements [5]. In [16], we employed a sparsity-based estimation technique to capture color and NIR images (four-channel images) using the Bayer color filter array.

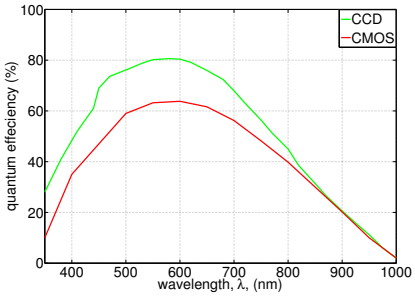


Figure 1: The quantum efficiency of typical CCD and CMOS sensors made of silicon. Both sensors are sensitive to visible (400-700 nm) and NIR (700-1100 nm) bands. Data for the CCD sensor is obtained from [15], and data for the CMOS sensor is provided by [11].

In the next section, we describe the sensor used for our experiments. We then explain our signal-separation algorithm. Finally, we present the experiments performed to evaluate our algorithm. The simulation results show a significant improvement in spectral accuracy of the data after applying our demultiplexing technique.

Multispectral Sensor Characteristics

Our multispectral imaging prototype uses a commercial silicon sensor. A layer of transmittance filters with different characteristics is mounted on the sensor. The filters are arranged according to the SFA proposed by Miao *et al.* [12], shown in Figure 2. The system is driven by an electronic board and a Zedboard that permits communication, and a software that enables capturing both images and video sequences. The sensor provides images with eight spectral bands and spatial resolution of 320 $\times$ 256 pixels.

A spectral calibration performed by a monochromator provides the relative sensor response in the range of $380-1100$ nm with steps of 10 nm [19]. The spectral sensitivities are shown in Figure 3. In the following, we refer to this sensor as the *real* sensor. We number the filters (and their corresponding spectral channels) in the order of their peak transmittance from 1 to 8,

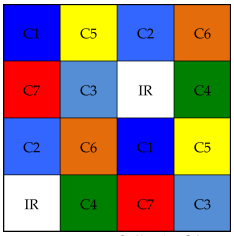


Figure 2: Spatial arrangement of SFA filters in our system. The SFA is designed by the binary tree algorithm of [12]. Figure 3 shows the spectral characteristics of filters in our prototype.

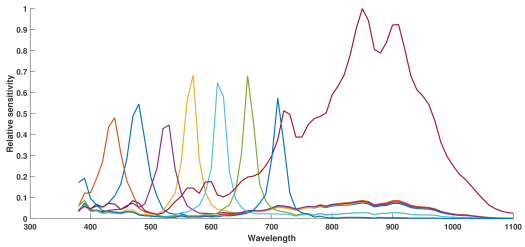


Figure 3: Spectral sensitivities of our multispectral sensor. Sensitivities from left to right correspond to channels 1 to 8.

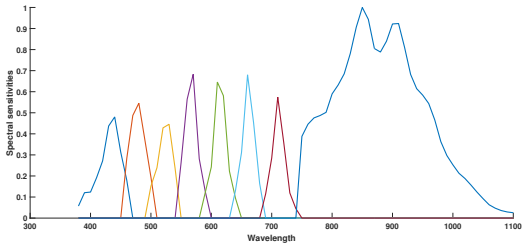


Figure 4: Spectral sensitivities of an ideal multispectral sensor. These sensitivities are not physically realizable. The curves are used solely for comparison.

where filter 1 transmits mainly the blue light, and filter 8 is the NIR filter.

As shown in Figure 3, visible filters transmit NIR radiation and the NIR filter transmits visible light. This leads to the sensor measurements being a mixture of visible and NIR information in each pixel. Hence, this sensor does not capture an accurate spectral signature of the scene.

An arbitrary ideal situation, which usually cannot be achieved in practice due to physical limitations, is when filters transmit the light only in a narrow band around their peak transmittance. We illustrate the spectral sensitivities for one such system in Figure 4. We simulated these curves simply by setting the transmittance of our filters outside their spectral range to zero. In what follows, we call this sensor the *ideal* sensor.

In the next section, we present a post-processing technique that demultiplexes visible and NIR information captured by the real sensor (Fig. 3). By using only the mixed and subsampled sensor measurements, our technique estimates an accurate multispectral representation of the scene. Simulation results show that the output of our algorithm is a close approximation of the multispectral data captured by the ideal yet non-realizable sensor of Figure 4.

Demultiplexing Visible and NIR Channels

As shown in Figure 3, the NIR transmission of visible filters 5-7 in our imaging system is negligible. Hence, we can reasonably assume that the sensor receives only the visible radiation at the locations of these filters. Based on this observation, we develop an algorithm that separates the visible channels 1-4 and the NIR channel starting with the mixed sensor measurements.

Let us call the vector containing the intensities of four mosaiced visible channels and the NIR channel, as they have been captured by the ideal sensor of Figure 4, $\mathbf{x}$. $\mathbf{y}$ holds the mixed sensor measurements in the same pixels, as captured by the real sensor of Figure 3. The goal of our algorithm is to estimate $\mathbf{x}$ given $\mathbf{y}$. This estimation, similar to the general problem of signal separation, is underconstrained. To constrain the problem, we assume that multispectral images can be compactly represented in a transform domain trained to express the correlation in this class of signals.

To estimate the intensities ($\mathbf{x}$) in each local patch of the image, we first compute the sparse coefficients of spectral channels in the transform domain:

$$\hat{\mathbf{s}} = \text{argmin } \|\mathbf{s}\|_0 \quad \text{s.t. } \mathbf{y} = \mathbf{ADs}, \tag{1}$$

where $\mathbf{s}$ is the sparse representation of the data, with minimum ℓ_0 norm that indicates sparsity. $\mathbf{A}$ is the transform that maps the ideal spectral sensitivities shown in Figure 4 to those of our real sensor (Fig. 3). $\mathbf{D}$ is a transformation matrix that is designed to decorrelate the multispectral data in the wavelength range covered by our imaging system. To design $\mathbf{D}$, we apply the training algorithm of Aharon *et al.* [1] to ground-truth data simulated using the ideal spectral sensitivities (Fig. 4).

Once the sparse coefficients are computed by solving (1), the target signal $\mathbf{x}$ is estimated as $\hat{\mathbf{x}} = \mathbf{D}\hat{\mathbf{s}}$. This process is repeated for all the overlapping patches in the image to reconstruct visible channels 1-4 and the NIR channel.

The algorithm explained above removes the contribution of visible channels 1 to 4 from the measurements at the location of the NIR filter. However, the NIR filter transmits the radiation in the range of 600-700 nm as well. To correct for this, we interpolate mosaiced channels 5-7 and obtain their intensities in the location of the NIR filter. We then subtract the intensities of channels 5-7 from the NIR channel estimated in the previous step of the algorithm. In this step, we take the relative response of the sensor in the corresponding channels into account.

Experiments

In this section, we analyze the performance of our algorithm both for synthetic data and real multispectral data captured by the imaging system described previously.

Implementation Details

Our algorithm processes the patches of size 8×8 where the SFA size is 4×4. To increase the reconstruction quality, we use an overlap of six pixels in each dimension between sequential patches. To solve (1), the sparse decomposition algorithm of Mohimani *et al.* [13] is employed. After demultiplexing NIR and visible information, we apply the binary-tree demosaicing algorithm of Miao *et al.* [12] to obtain full-resolution multispectral images.

Table 1: The mean square error (MSE) values for the spectral data simulated using the sensitivity of the real sensor (labeled "Real-sensor data") and the output of our algorithm (called "*Ours*"). The errors are computed with respect to the data simulated by the ideal sensor.

Image	case	**Band 1**	**Band 2**	**Band 3**	**Band 4**	Band 5	Band 6	Band 7	**Band 8**
San Francisco	Real-sensor data	0.2076	0.0914	0.1301	0.0584	0.0103	0.0152	0.0069	0.0064
	Ours	0.0107	0.0108	0.0101	0.0054	0.0103	0.0152	0.0069	0.0001
Stanford Dish	Real-sensor data	0.3856	0.2133	0.3814	0.2125	0.0483	0.0451	0.0361	0.0188
	Ours	0.0282	0.0342	0.0428	0.0270	0.0483	0.0451	0.0361	0.0002
Stanford Tower	Real-sensor data	0.1855	0.0967	0.1599	0.0722	0.0118	0.0149	0.0049	0.0056
	Ours	0.0064	0.0084	0.0093	0.0047	0.0118	0.0149	0.0049	0.0001

Synthetic Data

To objectively evaluate the proposed algorithm, we use the hyperspectral radiance data of the SCIEN database [17]. These images contain spectral information in the range of 415 to 950nm.

Starting from the SCIEN hyperspectral data, we simulate the multispectral acquisition both by the real sensor and by the ideal one. The acquisition is simulated according to the following image formation model:

$$\rho_i = \int_\lambda R(\lambda) S_i(\lambda) d\lambda, \tag{2}$$

where ρ_i, the pixel intensity in channel i is obtained by integrating the product of the radiance image $R(\lambda)$ and the sensor sensitivity $S(\lambda)$ in channel i. To perform the above simulation, we down-sampled the image radiance data in the SCIEN dataset by linear interpolation to achieve the spectral sampling rate of 10 nm. We limited the sensor range to $[415, 950]$ nm.

We simulated multispectral acquisition by the real and ideal sensors for three images provided in the SCIEN database, namely *San Francisco*, *Stanford Dish*, and *Stanford Tower*. We then processed the image produced using the sensitivities of the real sensor by our algorithm to obtain a corrected image that estimates the ideal measurements. Figure 5 shows the color representations of these three scenes.

To evaluate the quality of results, we compute the mean square error (MSE) between the ideal measurements and those simulated using the real sensor sensitivities, and between the ideal measurements and the output of our algorithm. The MSE results are reported in Table 1, when the pixel intensities are normalized to the range of $[0, 1]$.

As explained before our algorithm does not process visible channels 5-7, as the NIR transmittance of the corresponding filters is negligible. As a result the MSE values of the real measurements and our results for these channels are exactly the same. Our algorithm, however, processes visible channels 1-4 and the NIR channel (channel 8) for which we observe significant improvements in terms of MSE compared to real-sensor measurements. Using our technique, the error for these channels is reduced to 1/20-1/60 of the error for the real-sensor measurements.

Figures 6-7 show crops of *San Francisco* and *Stanford dish* scenes[1] in visible channels 1-4 and the NIR channel. For each scene, we show three images: the ground-truth image simulated by the ideal sensor, the image simulated by the real sensor, and the result of our algorithm. These figures demonstrate that as opposed to real-sensor measurements, the intensities in our results closely approximate the ground-truth images. For instance, consider the vegetation in these figures. Vegetation reflects NIR more than the visible light and appears brighter in the NIR image. However, as visible filters of the real sensor transmit NIR radiation, vegetation looks much brighter in the visible channels of the real images. As shown in Figures 6-7, our algorithm successfully corrects real-sensor images, so the intensities in our results are closer to the ground-truth images.

Note that for each scene, the spectral channels are mosaiced according to the SFA in Figure 2, where the sampling rate of each channel is only 1/8. As a result, the spatial resolution of the demosaiced images is lower than the images captured by a color camera with the same sensor resolution. In fact, any single-sensor multispectral imager trades off spatial resolution to achieve a much higher spectral resolution compared with a color camera.

Real Data

This subsection presents the results of applying the proposed algorithm on real multispectral data captured by our imaging system. The objective of our algorithm is to increase the spectral accuracy of measured data. As no multispectral ground-truth is available for our real dataset[2], to evaluate the spectral accuracy, we form the sRGB image for each multispectral image.

Figure 8 shows the impact of our correction algorithm on the color acuity of the captured data. Figure 8-(a) is captured by a color camera. To compute the color image in Figure 8-(b) we applied the demosaicing algorithm of Miao *et al.* [12] and mapped the 8-band multispectral image into the sRGB color space using the spectral sensitivities of our sensor, XYZ color matching functions, and XYZ to sRGB color transformation. Figure 8-(c) is obtained with the same process except that before demosaicing, our algorithm is applied to separate mixed sensor measurements. As shown in this example, our algorithm results in a more accurate color reproduction. For instance, see the color of sky in the upper right part of the image, which appears greenish in Figure 8-(b), while it is blue both in the ground-truth color image and the result of our algorithm. The color cast in Figure 8-(b) might be caused by the higher NIR transmittance of the mid-wavelength filters in our imaging system compared to the filters transmitting the red light.

[1] Due to space constraints we omit the subjective comparisons for the *Stanford dish* scene. As shown in Table 1, similar improvements are obtained for this scene.

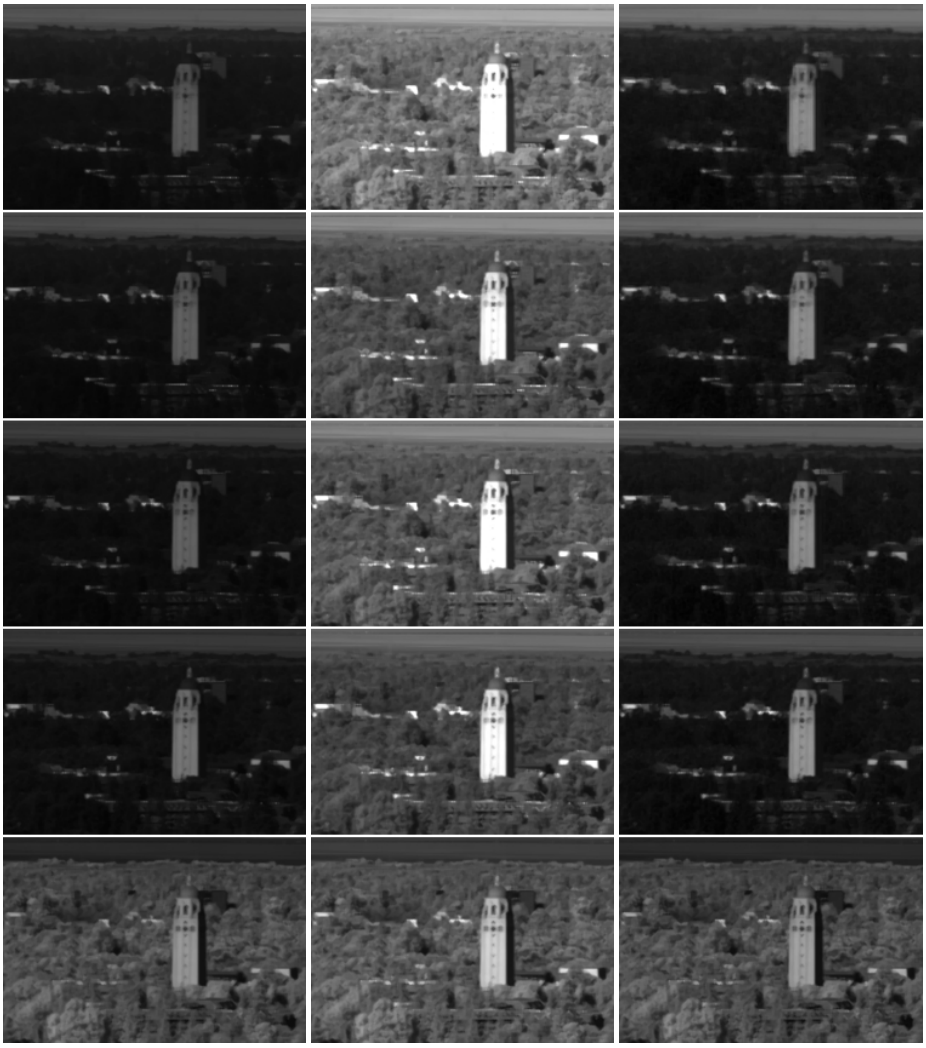
[2] Capturing ground-truth for this dataset requires an ideal multispectral imaging system in which the spectral transmittance of different filters do not overlap.

(a) The San Francisco Scene (b) The Stanford Dish Scene (c) The Stanford Tower Scene

Figure 5: The color representations of three scenes in the SCIEN dataset used in our experiments.

(a) The ground-truth images (b) The real-sensor images (c) *Our results*

Figure 6: A crop of the San Francisco scene from the SCIEN dataset. (a) The spectral images acquired by the ideal sensor (ground-truth images). (b) The spectral data simulated using the real sensor model. (c) Our results. From top to bottom: Visible channels 1-4 and the NIR channel (channel 8). Visible filters of the real sensor transmit the NIR radiation as well. As a result, vegetation appears much brighter in visible channels simulated based on the real sensor sensitivity. Moreover, the NIR filter transmits a portion of the visible light, hence the sky appears brighter in channel 8 of the real-sensor data. Our algorithm corrects for these issues by accurately estimating the visible and NIR intensities. The differences between pixel intensities are most noticeable when this figures is displayed on the screen.



(a) The ground-truth images (b) The real images (c) *Our results*

Figure 7: A crop of the Stanford Tower scene from the SCIEN dataset. (a) The spectral images acquired by the ideal sensor (ground-truth images). (b) The spectral data simulated using the real sensor model. (c) Our results. From top to bottom: Visible channels 1-4 and the NIR channel (channel 8). Our algorithm (c) corrects for the spectral inaccuracies observed in real-sensor images (b) compared with ground-truth (a). The differences between pixel intensities are most noticeable when this figures is displayed on the screen.

(a) Ground-truth color image (b) Without spectral separation (c) Our result

Figure 8: (a) The ground-truth image captured by a three-channel color camera. (b) The color image of the multispectral data captured by our imaging system. (c) The color image of the multispectral data corrected by our algorithm. Our algorithm results in a more accurate color reproduction of the scene. Note that the sensor resolution of the color camera is higher than our multispectral sensor.

Conclusion

A single-sensor imaging system with a multispectral filter array offers a more efficient solution in terms of price and convenience of use compared with conventional multispectral acquisition techniques. The transmittance precision of the manufactured filters used in such a system is, however, limited, resulting in strong leakage of other spectral bands into each filter. In this paper, we demonstrated this issue in our multispectral visible and NIR acquisition prototype and presented a post-processing technique that corrects for the leakage of the visible radiation into NIR pixels and vice versa.

As proposed in the source separation literature, we designed a transform domain that represents the spatial and spectral correlations for ground-truth multispectral data. Our algorithm demultiplexes NIR and visible information by projecting the mixed sensor measurements into this transform domain.

We analyzed the performance of our algorithm for the images of the SCIEN dataset, and showed that it provides a very close approximation of the data simulated by the ideal yet impractical multispectral sensor. The same improvement in the spectral accuracy was demonstrated for real data as well. These results ensure that our post-processing technique can overcome the inherent limitations in manufacturing spectral filters and provide an accurate spectral representation of the scene.

Acknowledgments

This work was supported by the BPI-France funding of the OFS project and the EU-H2020 project EXIST. We thank Dr. Pierre-Jean Lapray for his help in capturing the data used in this paper.

References

[1] M. Aharon, M. Elad, and A.M. Bruckstein. The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representations. *IEEE Transactions on Image Processing*, 54(11):4311–4322, 2006.

[2] P. Bofill and M. Zibulevsky. Underdetermined blind source separation using sparse representations. *Signal Processing*, 81(11):2353 – 2362, 2001.

[3] M. F. Duarte and R. G. Baraniuk. Kronecker compressive sensing. *IEEE Transactions on Image Processing*, 21(2):494–504, 2012.

[4] L. Frey, L. Masarotto, M. Armand, M. Charles, and O. Lartigue. Multispectral interference filter arrays with compensation of angular dependence or extended spectral range. *Opt. Express*, 23(9):11799–11812, 2015.

[5] M. Golbabaee, S. Arberet, and P. Vandergheynst. Compressive source separation: Theory and methods for hyperspectral imaging. *IEEE Transactions on Image Processing*, 22(12):5096–5110, 2013.

[6] J. Hershey and Z. Zhang. Multispectral digital camera employing both visible light and non-visible light sensing on a single image sensor. US patent, 2008.

[7] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi. Simultaneous capturing of RGB and additional band images using hybrid color filter array. *Proc. of IS&T/SPIE Electronic Imaging: Digital Photography X*, 2014.

[8] P. Lapray, J. Thomas, and P. Gouton. A multispectral acquisition system using MSFAs. *IS&T Color and Imaging Conference*, 2014.

[9] P. Lapray, X. Wang, J. Thomas, and P. Gouton. Multispectral filter arrays: Recent advances and practical implementation. *Sensors*, 14(11):21626–21659, 2014.

[10] Y.M. Lu, C. Fredembach, M. Vetterli, and S. Süsstrunk. Designing color filter arrays for the joint capture of visible and near-infrared images. *IEEE International Conference on Image Processing*, 2009.

[11] P. Magnan. Detection of visible photons in CCD and CMOS: A comparative view. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 504(1-3):199–212, 2003.

[12] L. Miao, H. Qi, R. Ramanath, and W.E. Snyder. Binary tree-based generic demosaicking algorithm for multispectral filter arrays. *IEEE Transactions on Image Processing*, 15(11):3550–3558, 2006.

[13] H. Mohimani, M. Babaie-Zadeh, and C. Jutten. A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm. *IEEE Transactions on Signal Processing*, 57(1):289–301, 2009.

[14] H. Park and K. B. Crozier. Multispectral imaging with vertical silicon nanowires. *Scientific reports*, 3(2460), 2013.

[15] M. Raffel, C.E. Willert, S.T. Wereley, and J. Kompenhans. *Particle Image Velocimetry*. Springer, 2007.

[16] Z. Sadeghipoor, Y.M. Lu, and S. Süsstrunk. A novel compressive sensing approach to simultaneously acquire color and near-infrared images on a single sensor. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2013.

[17] T. Skauli and J. Farrell. A collection of hyperspectral images for imaging systems research. *Proc. SPIE, Digital Photography IX*, 2013.

[18] J. L. Starck, M. Elad, and D. L. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Transactions on Image Processing*, 14(10):1570–1582, 2005.

[19] J. Thomas, P. Lapray, P. Gouton, and C. Clerc. Spectral characterization of a prototype sfa camera for joint visible and nir acquisition. *Sensors*, 16(7):993, 2016.