Convolutional Sparse Coding for RGB+NIR Imaging

Xuemei Hu, Felix Heide, Qionghai Dai, and Gordon Wetzstein

***Abstract*— Emerging sensor designs increasingly rely on novel color filter arrays (CFAs) to sample the incident spectrum in unconventional ways. In particular, capturing a near-infrared (NIR) channel along with conventional RGB color is an exciting new imaging modality. RGB+NIR sensing has broad applications in computational photography, such as low-light denoising, it has applications in computer vision, such as facial recognition and tracking, and it paves the way toward low-cost single-sensor RGB and depth imaging using structured illumination. However, cost-effective commercial CFAs suffer from severe spectral cross talk. This cross talk represents a major challenge in high-quality RGB+NIR imaging, rendering existing spatially multiplexed sensor designs impractical. In this work, we introduce a new approach to RGB+NIR image reconstruction using learned convolutional sparse priors. We demonstrate high-quality color and NIR imaging for challenging scenes, even including high-frequency structured NIR illumination. The effectiveness of the proposed method is validated on a large data set of experimental captures, and simulated benchmark results which demonstrate that this work achieves unprecedented reconstruction quality.**

***Index Terms*— Computational photography, convolutional sparse coding, structured illumination.**

I. Introduction

IMAGING in the near-infrared (NIR) spectral range is emerging as an exciting low-cost imaging modality beyond traditional RGB color imaging and has broad applications in physical and biological sciences [1], in computer vision, such as depth-imaging, feature detection [2], descattering [3], or dehazing [4], and in computational photography [5].

Manuscript received August 9, 2016; revised July 8, 2017 and September 30, 2017; accepted November 21, 2017. Date of publication December 8, 2017; date of current version January 5, 2018. This work was supported in part by the National Science Foundation of China under Grant 61327902 and Grant 61631009, in part by a Four-year Fellowship from The University of British Columbia, in part by the National Science Foundation under Grant IIS 1553333, in part by NSF/Intel Partnership on Visual and Experiential Computing under Grant NSF IIS 1539120, and in part by the Intel Compressive Sensing Alliance. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Dipti Prasad Mukherjee. *(Corresponding author: Qionghai Dai.)*

X. Hu is with the Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: huxm13@mails.tsinghua.edu.cn).

F. Heide and G. Wetzstein are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94306 USA (e-mail: fheide@stanford.edu; gordon.wetzstein@stanford.edu).

Q. Dai is with the Department of Automation, Tsinghua University, Beijing 100084, China, and also with the Zhejiang Future Technology Institute, Jiaxing 314006, China (e-mail: qhdai@tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2017.2781303

Unlike the human visual system, most silicon-based solid state sensors are sensitive to the NIR spectral region, and hence images in the NIR wavelength range can be acquired using mass-market CMOS sensors in combination with an optical bandpass filter suppressing the visible spectrum. Hence, as separate sensors, one for RGB color and one or more for NIR, RGB+NIR imaging systems can already be found in consumer products, such as Microsoft's Kinect and the Intel RealSense depth camera.[1] Recently, novel color filter arrays (CFAs) have emerged which multiplex RGB color channels as well as the NIR channel on the same sensor. These single-sensor RGB+NIR imagers make it possible to eliminate parallax, reduce bill of material and power requirements (by eliminating multiple sensors), and they have started to appear in experimental hardware platforms, such as Google's Project Tango.[2]

While these sensors, such as the OmniVision 4682 model, represent a promising avenue to efficient, low-cost devices, acquiring high-quality images with RGB+NIR CFAs remains a challenging research problem which currently prohibits broad practical applications. RGB color filters are selectively transmissive in the visible spectrum but usually transparent in the NIR domain, and hence RGB color filters receive significant crosstalk from the NIR channel. This results in RGB colors appearing "washed out" in the presence of NIR light, which is why consumer color cameras rely on an IR cutoff filter. While such washed-out images may still be acceptable for some computer vision applications, structured NIR illumination, such as the speckle pattern used in the Intel RealSense, causes severe localized artefacts. In addition to corrupted RGB color, the NIR channel is usually implemented by omitting a color filter in front of a pixel so that it becomes panchromatic. Currently, commercial RGB+NIR sensors are used in combination with an optical NIR blocker to capture "clean" RGB channels and an additional panchromatic channel in the visible spectrum, an imaging mode referred to as RGBW. Hence, one may conceive a sequential capture mode, with and without an IR cutoff filter, as a potential way to separate NIR and RGB information without introducing parallax. However, this approach necessitates solving major optical challenges as it requires two alternate optical paths that are dynamically changed, or mechanical moving filters, which would increase cost, form factor, and frame latency.

[1] https://software.intel.com/en-us/realsense/home

[2] https://get.google.com/tango/

To eliminate crosstalk in RGB+NIR imaging, Tang *et al.* [6] recently introduced an image reconstruction framework that attempts to recover unmixed RGB+NIR images as an inverse problem from only the spectrally mixed CFA measurements. To tackle the challenging ill-posed inverse problem, they propose a maximum a posteriori estimation, assuming Gaussian noise and a total variation (TV) image prior. Recognizing the essential role of the image prior for this spectral unmixing problem, the approach proposed in this work relies on learned, natural image statistics for this task. In particular, we adopt an optimization framework using convolutional sparse coding (CSC) for RGB+NIR image reconstruction. CSC models each individual color channel as a sparse sum of image patches (i.e. atoms) that are learned from training data and stored in a convolutional dictionary. We demonstrate that the proposed approach achieves high-quality image reconstructions, outperforming all existing approaches by a substantial margin in both simulation and experimental results. We are the first to demonstrate single-image RGB+NIR reconstruction with high-frequency structured NIR illumination patterns, paving the way towards low-cost single-sensor RGBD cameras.

Specifically, the key contributions of this work are:

- We reframe the RGB+NIR reconstruction problem as a convolutional sparse coding problem, allowing us to incorporate learned, natural images representations that address the ill-posedness of the underlying inverse problem.
- We develop a new and efficient reconstruction algorithm based on the alternating direction method of multipliers algorithm.
- We validate our approach on a large test set of representative simulated and captured RGB+NIR scenarios, with and without structured NIR illumination, and verify its performance compared to competing state-of-the-art approaches. The proposed method outperforms existing methods in a wide range of simulated and real-world scenarios.

In the following sections, we first review related work in Sec. II. In Sec. III, we introduce the image formation model and the proposed convolutional sparse coding framework. Next, in Sec. IV, the proposed method is validated in simulation and using experimental measurements. Finally, in Sec. V, we discuss the limitations and potential future directions of research building on this work.

II. Related Work

A. RGB+NIR Imaging

Recent approaches in NIR imaging extend the basic RGB mosaic with a fourth filter type with high transmittance in the NIR spectrum band. Using such a custom CFA, RGB+NIR information can be captured with a single-sensor image, but at the cost of reduced image resolution in the green channel compared to the traditional Bayer CFA pattern. [7] proposes an optimized CFA design for effective RGB and NIR image capture. [8] and [9] study demosaicing and crosstalk in isolation. [10] proposes compressive sensing for the recovery of RGB+NIR imaging. Note that this approach only considers the crosstalk between the green and NIR channel. Recently, [6] proposed an RGB+NIR demosaicing algorithm that models crosstalk and the defocus of each spectral channel. Although the authors of [6] solve a challenging inverse reconstruction problem, they rely on engineered, handcrafted quadratic or gradient sparsity regularizers. In contrast, this work relies on learned image priors to facilitate high-quality reconstruction of RGB and NIR channels. Having discussed RGB+NIR imaging under general illumination conditions, we next discuss related work relying on structured illumination.

B. Structured Illumination

Structured NIR illumination has been of interest to the research community for about a decade [11] and has mostly been used for stereo depth imaging. Speckle decorrelation [12] and depth-varying light field [13] methods have been proposed as structured illumination patterns for 3D ranging applications. Structured illumination has matured as a technology and is available in a range of consumer RGBD cameras, such as Microsoft's Kinect and the Intel RealSense. These devices have enabled a broad range of exciting applications in consumer electronics, robotics, machine vision, and beyond [14], [15]. Conventional RGBD cameras are equipped with two sensors, an RGB sensor and a separate NIR sensor. If a high-quality single-sensor RGB+NIR solution did exist, one could acquire a single image to recover a high-quality RGB image along with an NIR channel that contains the structured illumination. Such an approach would reduce not only the form factor, cost, and power of RGBD cameras but would also remove the need for registration between the two cameras. Unfortunately, no image processing technique is currently capable of actually recovering high-quality RGB images when the NIR structured illumination contaminates the measurements of the color channels. In this paper, we demonstrate the first approach to achieve high-quality RGB imaging with structured NIR illumination. Reconstructing high-quality RGB and NIR images can also be seen as a demosaicing problem with an unconventional color filter array. In the following subsection, we review the related work on demosaicing.

C. Image Demosaicing

Demosaicing, the inpainting of spatially subsampled spectral measurements, is a mature field with an immense body of prior work [16]. Both CFAs and corresponding demosaicing algorithms have been extensively optimized for RGB color imaging [17]. Demosaicing can be divided into two branches, one considering spatial-domain reconstructions and the other considering frequency-domain reconstructions [18]. Many variants of CFAs have been proposed, as in [19] and [20], including several versions of RGB+NIR CFAs. Due to manufacturing limitations, crosstalk among the RGB and NIR channels is severe and needs to be modeled accurately for high-quality color reconstruction. Several methods have been proposed to demosaic and unmix measurements from such CFAs [6], [8], [10], [21], [22]. Due to the spectral unmixing

component, high-quality RGB+NIR demosaicing remains a challenging area of very active research. This work poses the underlying reconstruction problem as a convolutional sparse coding problem that is solved using optimization. The next section provides a brief review of related optimization methods in imaging.

D. Image Optimization

Research in the field of natural image statistics suggests that natural images contain statistical structures that set them apart from purely random signals [23]. Therefore, characterizing the inherent structure of natural images and formulating efficient representations based on these structures provides essential insights into the recovery of natural images [24], in the form of image priors under a Bayesian model. Sparse coding algorithms have been proposed for learning image patches as the basic structures of natural images [24], [25]. Each patch can be represented by a sparse linear combination of the learned atoms [26], [27]. Beyond 2D image processing, patch-based sparse coding has also been applied to 3D hyperspectral imaging [28] and 4D light field imaging [29]. Unfortunately, patch-based representations ignore the spatial invariance of images. Shifted versions of the same image patch have to be represented, causing patch-based dictionaries to be highly redundant. Convolutional sparse coding has been proposed as an alternative to patch-based sparse coding, which resolves this redundancy [30]–[32]. In contrast to patch-based methods, CSC operates on whole images, thereby seamlessly capturing the correlation between local neighborhoods. CSC was introduced in the context of modeling receptive fields in human vision [33]. It has since been demonstrated to have applications in a wide range of computer vision problems, such as low/mid-level feature learning [34], [35], image restoration [36], [37], and computational imaging problems [38]–[41]. In this work, we learn convolutional sparse representations for RGB+NIR images and rely on these as efficient priors resolving the ill-posedness of RGB+NIR image reconstruction.

III. Convolutional Sparse Demosaicing, Deconvolution, and Spectral Unmixing

In this section, we describe the proposed reconstruction method. First, we briefly introduce the image formation model for single-sensor RGB+NIR imaging. Next, the convolutional sparse representation and learning of the proposed image prior are described. We then outline the proposed image reconstruction method and finally provide in-depth derivations for the reconstruction algorithm and calibration methods.

A. Image Formation Model

We adopt the image formation model proposed by Tang *et al.* [6]. For completeness, we briefly review this model before deriving a novel reconstruction framework for estimating RGB+NIR channels from a single sensor image. Following [6], we assume that the camera measurements do not saturate, i.e. they are not under- or overexposed. This could be handled by an additional masking operator [41] that effectively inpaints saturated regions. We denote a 2D measurement image as $\boldsymbol{J}$ and its vectorized form as $\boldsymbol{j}$, which can be modeled as

$$\boldsymbol{j} = \sum_{i\in\{\mathrm{R,G,B,I}\}} \boldsymbol{S}_i \left(\int r_i(\lambda) \boldsymbol{K}_\lambda \boldsymbol{l}_\lambda \, \mathrm{d}\,\lambda \right) + \boldsymbol{n}, \tag{1}$$

where all images are represented by column vectors. The latent full resolution image for a given wavelength λ is denoted as $\boldsymbol{l}_\lambda \in \mathbb{R}^N$, with N representing the total number of pixels in the image. $\boldsymbol{K}_\lambda \in \mathbb{R}^{N\times N}$ is the convolution matrix modeling wavelength-dependent blur in the imaging optics. The light incident at the sensor is multiplied with $r_i\ (\lambda) \in \mathbb{R}$, which is the spectral response of color filter i ($i \in \{\mathrm{R,G,B,I}\}$) at wavelength λ, as shown in Fig. 1(a). We use $\mathrm{R,G,B,I}$ to denote the red, green, blue, and NIR color channel; $\boldsymbol{n}$ denotes additive sensor noise. Finally, $\boldsymbol{S}_i \in \mathbb{R}^{N\times N}$ describes the subsampling operator corresponding to color channel i and is a diagonal matrix. Defining this operator as

$$\boldsymbol{S}_i[t,t] = \begin{cases} 1 & \text{if pixel at position } t \text{ has color filter } i, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

allows modeling the captured image $\boldsymbol{j}$ in Eq. (1) as a sum of four linear transforms. Here, the index t denotes a coefficient location in the vector $\boldsymbol{j}$, which corresponds to the position of the pixel in $\boldsymbol{J}$. The ideal color filter for $\mathrm{R,G,B}$ would only have spectral support in the respective visible wavelength range and the NIR color filter to only have spectral support in the NIR range. In particular, each ideal color filter would have a spectral response concentrated

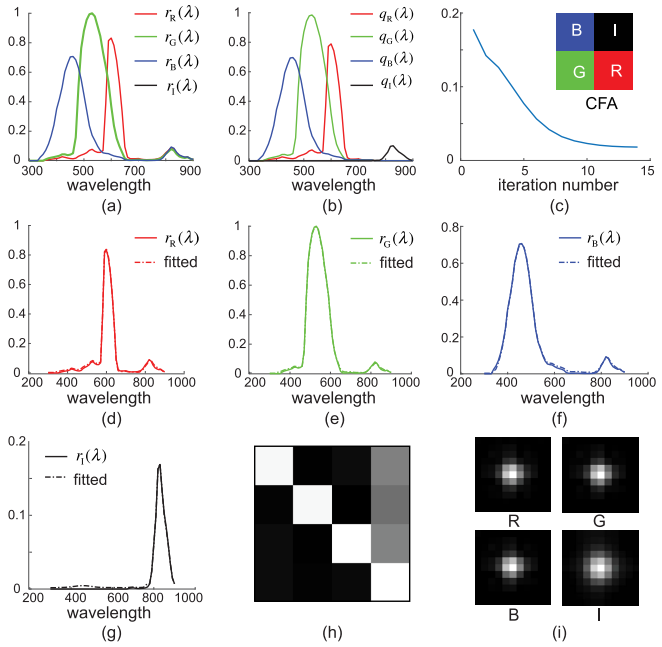


Fig. 1. (a) Measured spectral response of the OmniVision OV4682 sensor. (b) Optimized, ideal narrow-band spectral responses. Note that the crosstalk in the IR range is eliminated. (c) Convergence plot of the crosstalk matrix estimation. (d)-(g) are the R, G, B, I spectral response $\boldsymbol{R}$, and the fitted results $\boldsymbol{C} * \boldsymbol{Q}$ at the estimated crosstalk matrix $\boldsymbol{C}$. (h) Estimated crosstalk matrix $\boldsymbol{C}$, corresponding to (b). (i) Calibrated point spread functions for all channels.

around a peak spectral response, only covering a narrow band-pass range in the visible-to-NIR wavelength range, as shown in Fig. 1(b). However, manufacturing constraints prevent the filters of a color mosaic from being ideal narrow band-pass filters. The measured filter responses in Fig. 1(a) show that the spectral response of the R, G, B cause crosstalk in the NIR range. Assuming that the color spectrum $r(\lambda) = [r_{\mathrm{R}}(\lambda), r_{\mathrm{G}}(\lambda), r_{\mathrm{B}}(\lambda), r_{\mathrm{I}}(\lambda)]^{\mathrm{T}}$, which is a 4×1 column vector, it is $q(\lambda) = [q_{\mathrm{R}}(\lambda), q_{\mathrm{G}}(\lambda), q_{\mathrm{B}}(\lambda), q_{\mathrm{I}}(\lambda)]^{\mathrm{T}}$, and this crosstalk can be expressed as the 4×4 matrix $\boldsymbol{C}$

$$r(\lambda) = \boldsymbol{C} q(\lambda), \tag{3}$$

where

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_{\mathrm{RR}} & \boldsymbol{C}_{\mathrm{GR}} & \boldsymbol{C}_{\mathrm{BR}} & \boldsymbol{C}_{\mathrm{IR}} \\ \boldsymbol{C}_{\mathrm{RG}} & \boldsymbol{C}_{\mathrm{GG}} & \boldsymbol{C}_{\mathrm{BG}} & \boldsymbol{C}_{\mathrm{IG}} \\ \boldsymbol{C}_{\mathrm{RB}} & \boldsymbol{C}_{\mathrm{GB}} & \boldsymbol{C}_{\mathrm{BB}} & \boldsymbol{C}_{\mathrm{IB}} \\ \boldsymbol{C}_{\mathrm{RI}} & \boldsymbol{C}_{\mathrm{GI}} & \boldsymbol{C}_{\mathrm{BI}} & \boldsymbol{C}_{\mathrm{II}} \end{bmatrix}. \tag{4}$$

$\boldsymbol{C}$ models the superposition of the ideal narrow band-pass response to the actual color filter spectral response [6].

As observed in Fig. 1(a), the response of the red color filter r_{R}, for example, can be seen as the superposition of the ideal red spectral response q_{R} weighted by $\boldsymbol{C}_{\mathrm{RR}}$ and the ideal NIR spectral response q_{I} weighted by $\boldsymbol{C}_{\mathrm{IR}}$.

Eq. (3) and Eq. (1) yield

$$\boldsymbol{j} = \sum_{i \in \{\mathrm{R,G,B,I}\}} \boldsymbol{S}_i \left(\int \sum_{i' \in \{\mathrm{R,G,B,I}\}} \boldsymbol{C}_{i'i} q_{i'}(\lambda) \boldsymbol{K}_\lambda \boldsymbol{l}_\lambda \, \mathrm{d}\, \lambda \right) + \boldsymbol{n}. \tag{5}$$

For the ideal narrow band-pass color filters, the blur kernel of each color channel is assumed to be wavelength-independent, denoted by the convolution matrices $\boldsymbol{K}_{\mathrm{R}}$, $\boldsymbol{K}_{\mathrm{G}}$, $\boldsymbol{K}_{\mathrm{B}}$ and $\boldsymbol{K}_{\mathrm{I}}$. Now, inserting $\boldsymbol{K}_{\mathrm{R}}$, $\boldsymbol{K}_{\mathrm{G}}$, $\boldsymbol{K}_{\mathrm{B}}$ and $\boldsymbol{K}_{\mathrm{I}}$ into Eq. (5) and changing the integrand, we rewrite Eq. (5) as

$$\boldsymbol{j} = \sum_{i \in \{\mathrm{R,G,B,I}\}} \boldsymbol{S}_i \left(\sum_{i' \in \{\mathrm{R,G,B,I}\}} \boldsymbol{C}_{i'i} \boldsymbol{K}_{i'} \int q_{i'}(\lambda) \boldsymbol{l}_\lambda \, \mathrm{d}\, \lambda \right) + \boldsymbol{n}. \tag{6}$$

Since $q_{i'}(\lambda)$ is the ideal spectral response, the ideal image that we seek to recover is $\int q_{i'}(\lambda) \boldsymbol{l}_\lambda \, \mathrm{d}\, \lambda$. Denoting this ideal image as $\boldsymbol{h}_{i'}$, we can reformulate Eq. (6) as

$$\boldsymbol{j} = \sum_{i \in \{\mathrm{R,G,B,I}\}} \boldsymbol{S}_i \left(\sum_{i' \in \{\mathrm{R,G,B,I}\}} \boldsymbol{C}_{i'i} \boldsymbol{K}_{i'} \boldsymbol{h}_{i'} \right) + \boldsymbol{n}. \tag{7}$$

To get a compact image formation model, we stack the four latent variables $\boldsymbol{h}_R$, $\boldsymbol{h}_G$, $\boldsymbol{h}_B$, $\boldsymbol{h}_I$ into a single variables $\boldsymbol{h}$, that is $\boldsymbol{h} = \left[\boldsymbol{h}_{\mathrm{R}}^{\mathrm{T}}, \boldsymbol{h}_{\mathrm{G}}^{\mathrm{T}}, \boldsymbol{h}_{\mathrm{B}}^{\mathrm{T}}, \boldsymbol{h}_{\mathrm{I}}^{\mathrm{T}}\right]^{\mathrm{T}}$, and use $\boldsymbol{S} = [\boldsymbol{S}_{\mathrm{R}}, \boldsymbol{S}_{\mathrm{G}}, \boldsymbol{S}_{\mathrm{B}}, \boldsymbol{S}_{\mathrm{I}}]$ and $\boldsymbol{K} = \mathrm{diag}\,(\boldsymbol{K}_{\mathrm{R}}, \boldsymbol{K}_{\mathrm{G}}, \boldsymbol{K}_{\mathrm{B}}, \boldsymbol{K}_{\mathrm{I}})$, which changes Eq. (7) to

$$\boldsymbol{j} = \boldsymbol{S}(\boldsymbol{C} \otimes \mathbb{I}_N) \boldsymbol{K} \boldsymbol{h} + \boldsymbol{n}. \tag{8}$$

Here, the image formation operator is composed of the convolution matrix $\boldsymbol{K}_i$ that convolves each channel with the respective blur kernel, followed by a crosstalk matrix $\boldsymbol{C} \otimes \mathbb{I}_N$ that applies crosstalk mixing between the blurred channels, and finally a subsampling matrix $\boldsymbol{S}$. The convolution matrix $\boldsymbol{K}$ is a block-diagonal matrix that has the individual convolution matrices for each channel as diagonal elements.

Using the expression $\boldsymbol{F} = \boldsymbol{S}(\boldsymbol{C} \otimes \mathbb{I}_N)\boldsymbol{K}$, we can compactly formulate the image formation model as

$$\boldsymbol{j} = \boldsymbol{F}\boldsymbol{h} + \boldsymbol{n}. \tag{9}$$

All components of the image formation in Eq. (9) are assumed to be known for the inference of $\boldsymbol{h}$ from $\boldsymbol{j}$. The convolution matrix $\boldsymbol{K}$ and the crosstalk matrix $\boldsymbol{C}$ are estimated in a pre-calibration step, which we will discuss below in Sec. III-E, while $\boldsymbol{S}$ is known from the sensor layout, as defined in Eq. (2).

Given the image formation model from Eq. (9), the RGB+NIR imaging task that this paper addresses is to recover the ideal channel images $\boldsymbol{h}$ from the input measurement $\boldsymbol{j}$, which is an ill-posed joint demosaicing, spectral unmixing and deconvolution problem.

The ill-posedness of this problem can be resolved by adopting a Bayesian model that relies on image priors. To take advantage of the structure in natural images, we rely on convolutional sparse priors that recently have been shown to achieve high-quality results across a variety of imaging results [31], [41]. Next, we describe how we learn these image priors.

B. Convolutional Sparse Prior

CSC models a vectorized image $\boldsymbol{x} \in \mathbb{R}^N$ as a sum of sparsely distributed convolutional features [30], that is, $\boldsymbol{x}$ is modeled as

$$\boldsymbol{x} = \sum_{k=1}^{K} \boldsymbol{d}_k * z_k. \tag{10}$$

In other words, a convolutional representation for images is formed using the filter dictionary $\boldsymbol{d}_k \in \mathbb{R}^P$, $k \in \{1, \ldots, K\}$. The filter dictionaries $\boldsymbol{d}_k$ are capable of representing the underlying convolutional features of natural images. The convolutional codes are highly compressible, which means the corresponding feature maps $\boldsymbol{z}_k \in \mathbb{R}^n$, $k \in \{1, \ldots, K\}$ are learned to be sparse. The operator $*$ is the 2D convolutional operator defined on the vectorized inputs.

The filter dictionary is learned by solving the optimization problem over a large image database, that is

$$\begin{aligned} \underset{\boldsymbol{d}_k, \boldsymbol{z}_k^i}{\operatorname{argmin}} \quad & \sum_{i=1}^{n} \frac{1}{2} \left\| \boldsymbol{x}^i - \sum_{k=1}^{K} \boldsymbol{d}_k * \boldsymbol{z}_k^i \right\|_2^2 + \beta \sum_{i=1}^{n} \sum_{k=1}^{K} \left\| \boldsymbol{z}_k^i \right\|_1 \\ \text{subject to} \quad & \|\boldsymbol{d}_k\|_2^2 \leq 1 \quad \forall k \in \{1, \ldots, K\}, \end{aligned} \tag{11}$$

where $\boldsymbol{x}^i$ represents the training images, n is the number of training images, $\boldsymbol{z}_k^i$ is the corresponding feature map of $\boldsymbol{x}^i$, and β is a weight for balancing the first and second objective term. Given such a prior learned on an image database, we formulate the proposed image reconstruction method next.

C. Image Reconstruction

For the image formation model in Eq. (9), the corresponding inverse problem is to reconstruct $\boldsymbol{h}$ from $\boldsymbol{j}$, where $\boldsymbol{h}$ is an unknown $4N \times 1$ vector and $\boldsymbol{j}$ is a known $N \times 1$ measurement. In other words, this inverse problem is ill-posed. Adopting a

TABLE I
NOTATIONS USED IN THIS MANUSCRIPT

$\boldsymbol{J}, \boldsymbol{j}$	2D input image and its column-vectorized form
$\mathrm{R}, \mathrm{G}, \mathrm{B}, \mathrm{I}$	R, G, B, NIR color channel
$\boldsymbol{S}$	subsample matrix
r	calibrated spectral response
$\boldsymbol{R}$	column stacked matrix of r_{R}, r_{G}, r_{B}, r_{I}
λ	wavelength
$\boldsymbol{K}$	point spread function (PSF) matrix
$\boldsymbol{l}$	ground truth image of a spectral band
$\boldsymbol{n}$	additive sensor noise
t	index of pixel position
N	total pixel number
$\boldsymbol{C}$	crosstalk matrix
q	ideal narrow-band spectral response
$\boldsymbol{Q}$	column stacked matrix of q_{R}, q_{G}, q_{B}, q_{I}
$\boldsymbol{h}$	ideal narrow band image
$\boldsymbol{F}$	full image formation operator
$\boldsymbol{x}$	training image of CSC
$\boldsymbol{d}$	trained dictionary of CSC
$\boldsymbol{D}$	matrix operator of dictionary covolution
$\boldsymbol{z}$	sparse feature map of CSC
$\boldsymbol{o}$	offset term of CSC
β, γ	balancing parameters for prior, offset term
∇	gradient operator
$\boldsymbol{I}_p$	penality coefficient matrix
$\boldsymbol{y}, \boldsymbol{v}$	intermediate variables for ADMM
ρ, e	penalty parameters and updating step size in ADMM
$w, \boldsymbol{W}$	noise weighting coefficient and matrix
$u, u1, u2$	superscript for each 2×2 color filter mosaic
σ	standard deviation of the synthetic noise level
$\otimes$	Kronecker product
$\odot$	element-wise product

Bayesian approach, we use the learned convolutional sparse coding prior to address this ill-posedness.

Specifically, we first learn the image filter dictionary $\{\boldsymbol{d}_k\}$ through optimizing Eq. (11) (see [30]) from a database of natural images. Given the learned filter dictionary $\{\boldsymbol{d}_k\}$, we next formulate a convolutional sparse representation for the unknown latent image $\boldsymbol{h}_i$, yielding the convolutional reconstruction problem

$$\begin{aligned} &\underset{z_k^i}{\operatorname{argmin}} \quad \|\boldsymbol{j} - \boldsymbol{F}\boldsymbol{h}\|_2^2 + \beta \sum_i \sum_{k=1}^{K} \left\| z_k^i \right\|_1 \\ &\text{subject to } \boldsymbol{h}_i = \sum_{k=1}^{K} \boldsymbol{d}_k * z_k^i \quad \forall i \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}, \mathrm{I}\}. \end{aligned} \tag{12}$$

The constraints encode the convolutional sparse representation of $\boldsymbol{h}_i$, and z_k^i represent the filter map of $\boldsymbol{h}_i$ corresponding to filter $\boldsymbol{d}_k$.

However, directly incorporating the learned CSC filters into the reconstruction is not possible because the filter dictionary d_k^i is learned with whitened data using local contrast-normalization. The whitening removes offset and scaling in different image locations. While this processing is essential in the learning process, it prohibits the straightforward use of the dictionary as a generative model. While the correct scaling for recovery can be obtained during the optimization by finding the correct values in the sparse maps, the offset is not modeled in Eq. (12). To address this, we introduce an offset variable $\boldsymbol{o}$ as proposed in [41], yielding the following optimization problem

$$\begin{aligned} &\underset{z_k^i}{\operatorname{argmin}} \quad \|\boldsymbol{j} - \boldsymbol{F}\boldsymbol{h}\|_2^2 + \beta \sum_i \left(\sum_{k=1}^{K} \left\| z_k^i \right\|_1 + \gamma \, \|\nabla \boldsymbol{o}_i\|_2^2 \right) \\ &\text{subject to } \boldsymbol{h}_i = \sum_{k=1}^{K} \boldsymbol{d}_k * z_k^i + \boldsymbol{o}_i \quad \forall i \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}, \mathrm{I}\}. \end{aligned} \tag{13}$$

By expressing the smoothness term as the convolution $\boldsymbol{o}_i = \boldsymbol{d}_{K+1} * z_{K+1}^i$, with $\boldsymbol{d}_{K+1}$ being a Dirac delta function. As introduced in Table I, ∇ is the gradient operator. The problem can be written as

$$\begin{aligned} &\underset{z_k^i}{\operatorname{argmin}} \quad \|\boldsymbol{j} - \boldsymbol{F}\boldsymbol{h}\|_2^2 + \beta \sum_i \left(\sum_{k=1}^{K} \left\| z_k^i \right\|_1 + \gamma \left\| \nabla z_{K+1}^i \right\|_2^2 \right) \\ &\text{subject to } \boldsymbol{h}_i = \sum_{k=1}^{K+1} \boldsymbol{d}_k * z_k^i \quad \forall i \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}, \mathrm{I}\}. \end{aligned} \tag{14}$$

We observe that the noise differs for different color channels. Typically, the NIR channel is degraded more severely by noise than the RGB channels. Therefore, we add a weighting matrix in the data fidelity term as

$$\begin{aligned} &\underset{z_k^i}{\operatorname{argmin}} \quad \|\boldsymbol{W}(\boldsymbol{j} - \boldsymbol{F}\boldsymbol{h})\|_2^2 + \beta \sum_i \left(\sum_{k=1}^{K} \left\| z_k^i \right\|_1 + \gamma \left\| \nabla z_{K+1}^i \right\|_2^2 \right) \\ &\text{subject to } \boldsymbol{h}_i = \sum_{k=1}^{K+1} \boldsymbol{d}_k * z_k^i \quad \forall i \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}, \mathrm{I}\}, \end{aligned} \tag{15}$$

where $\boldsymbol{W}$ is a diagonal matrix balancing the reconstruction of the four channels. $\boldsymbol{W}$ is estimated in a pre-calculation step through noise estimation for each color channel. Eq. (15) is a convex optimization problem and here we use the Alternating Direction Method of Multipliers [42] algorithm (ADMM) to solve it. In the next subsection, we describe the reconstruction algorithm in detail.

D. ADMM Reconstruction Algorithm

This section describes the optimization algorithm we propose to solve Eq. (15). For notational simplicity, we define

$$\boldsymbol{z} = [\left(z_1^R\right)^T, \ldots, \left(z_{K+1}^R\right)^T, \ldots, \left(z_1^I\right)^T, \ldots, \left(z_{K+1}^I\right)^T]. \tag{16}$$

Furthermore, we introduce $\boldsymbol{y} = \boldsymbol{K}\boldsymbol{D}\boldsymbol{z}$ as a slack variable that allows to separately optimize all convolutional variables in the frequency domain and solve for the other variables in the spatial domain. The matrix $\boldsymbol{D}$ expresses the sum of the convolution with the filters, that is $\boldsymbol{D}\boldsymbol{z} = \sum_{k=1}^{K+1} \boldsymbol{d}_k * \boldsymbol{z}_k$. We also introduce $\boldsymbol{v} = \boldsymbol{z}$ as a slack variable for the sparsity constraint. With these substitutions, the problem from Eq. (15) becomes

$$\begin{aligned} &\underset{\boldsymbol{y}}{\operatorname{argmin}} \quad \|\boldsymbol{W}(\boldsymbol{j} - \boldsymbol{S}(\boldsymbol{C} \otimes \mathbb{I}_N)\boldsymbol{y})\|_2^2 \\ &\qquad\qquad + \beta \sum_{k=1}^{K} \|\boldsymbol{v}_k\|_1 + \gamma \, \|\nabla \boldsymbol{z}_{K+1}\|_2^2 \\ &\text{subject to } \boldsymbol{y} = \boldsymbol{K}\boldsymbol{D}\boldsymbol{z}, \quad \boldsymbol{v} = \boldsymbol{z}, \end{aligned} \tag{17}$$

Algorithm 1 ADMM Algorithm for Convolutional Sparse RGB+NIR Imaging

1: Initialize: $\rho = \rho_0$, $e = 1.01$, $\rho_{max} = 1e5$
2: **for** iter =1: M **do**
3: $\boldsymbol{z}^{k+1} = \underset{\boldsymbol{z}}{\operatorname{argmin}} \ \frac{\rho}{2} \left\| \boldsymbol{K}\boldsymbol{D}\boldsymbol{z} - \boldsymbol{y}^k + \boldsymbol{\lambda}_1^k \right\|_2^2 + \frac{\rho}{2} \left\| \boldsymbol{z} - \boldsymbol{v}^k + \boldsymbol{\lambda}_2^k \right\|_2^2 + \gamma \left\| \nabla \boldsymbol{z}_{K+1}^i \right\|_2^2$
4: $\boldsymbol{y}^{k+1} = \underset{\boldsymbol{y}}{\operatorname{argmin}} \ \| \boldsymbol{W} (\boldsymbol{j} - \boldsymbol{S}(\boldsymbol{C} \otimes \mathbb{I})\boldsymbol{y}) \|_2^2 + \frac{\rho}{2} \left\| \boldsymbol{K}\boldsymbol{D}\boldsymbol{z}^{k+1} - \boldsymbol{y} + \boldsymbol{\lambda}_1^k \right\|_2^2$
5: $\boldsymbol{v}^{k+1} = \underset{\boldsymbol{v}}{\operatorname{argmin}} \ \beta \| \boldsymbol{v} \|_1 + \frac{\rho}{2} \left\| \boldsymbol{z}^{k+1} - \boldsymbol{v} + \boldsymbol{\lambda}_2^k \right\|_2^2$
6: $\boldsymbol{\lambda}_1^{k+1} = \boldsymbol{\lambda}_1^k + (\boldsymbol{K}\boldsymbol{D}\boldsymbol{z}^{k+1} - \boldsymbol{y}^{k+1})$, $\boldsymbol{\lambda}_2^{k+1} = \boldsymbol{\lambda}_2^k + (\boldsymbol{z}^{k+1} - \boldsymbol{v}^{k+1})$
7: $\rho^{k+1} = min(\rho_{max}, \rho^k \times e)$
8: **end for**

where we solve only for the sparse feature maps $\boldsymbol{z}$. The reconstruction result is then $\boldsymbol{h} = \boldsymbol{D}\boldsymbol{z}$. This modified objective can be solved with ADMM algorithm given in Algorithm 1.

Note that in line 4, optimizing $\boldsymbol{y}$ is a large-scale quadratic problem. However, it is separable into groups aligned with the 2×2 CFA pattern. The small quadratic sub-problems can be solved using a fast direct method in parallel for all separable problems. Specifically, we first rewrite the problem as

$$\underset{\boldsymbol{y}}{\operatorname{argmin}} \ \| \boldsymbol{W}\boldsymbol{S}(\boldsymbol{C} \otimes \mathbb{I}_N)\boldsymbol{y} - \boldsymbol{b}_1 \|_2^2 + \frac{\rho}{2} \| \boldsymbol{y} - \boldsymbol{b}_2 \|_2^2, \quad (18)$$

where $\boldsymbol{b}_1 = \boldsymbol{W}\boldsymbol{j}$ and $\boldsymbol{b}_2 = \boldsymbol{K}\boldsymbol{D}\boldsymbol{z}^{k+1} + \boldsymbol{\lambda}_1^k$ are known variables. This large quadratic convex optimization can normally be solved by finding the zero point of the derivative of the objective function. However, it requires inverting a large matrix, which is computationally infeasible, especially when the image resolution N is high. This is a common problem with the image reconstruction work. Exploiting the structure of the forward image formation matrix, we can further reduce the computational cost of the large quadratic problem.

As the subsampling matrix $\boldsymbol{S}$ and the weighting matrix $\boldsymbol{W}$ are replicated for each 2×2 CFA unit and the crosstalk matrix $\boldsymbol{C}$ models the crosstalk of the NIR light to the RGB color channel within each CFA, this large quadratic problem is separable into each small color filter unit and can be solved efficiently as follows

$$\underset{\boldsymbol{y}^{\mathrm{u}}}{\operatorname{argmin}} \ \left\| \boldsymbol{W}^{\mathrm{u}}\boldsymbol{S}^{\mathrm{u}}(\boldsymbol{C}^{\mathrm{u}} \otimes \mathbb{I}_4)\boldsymbol{y}^{\mathrm{u}} - \boldsymbol{b}_1^{\mathrm{u}} \right\|_2^2 + \frac{\rho}{2} \left\| \boldsymbol{y}^{\mathrm{u}} - \boldsymbol{b}_2^{\mathrm{u}} \right\|_2^2, \quad (19)$$

assuming the "BGIR" CFA pattern, then

$$\boldsymbol{W}^{\mathrm{u}} = \operatorname{diag}\left(\boldsymbol{w}_b, \boldsymbol{w}_g, \boldsymbol{w}_i, \boldsymbol{w}_r\right) \quad (20)$$

$$\boldsymbol{S}^{\mathrm{u}} = \big[\operatorname{diag}(0,0,0,1), \operatorname{diag}(0,1,0,0), \operatorname{diag}(1,0,0,0), \operatorname{diag}(0,0,1,0)\big] \quad (21)$$

$$\boldsymbol{C}^{\mathrm{u}} = \boldsymbol{C} \otimes \mathbb{I}_4 \quad (22)$$

$$\boldsymbol{y}^{\mathrm{u}} = \Big[\boldsymbol{y}_{\mathrm{R}}^{u1,u1}, \boldsymbol{y}_{\mathrm{R}}^{u2,u1}, \boldsymbol{y}_{\mathrm{R}}^{u1,u2}, \boldsymbol{y}_{\mathrm{R}}^{u2,u2}, \boldsymbol{y}_{\mathrm{G}}^{u1,u1}, \boldsymbol{y}_{\mathrm{G}}^{u2,u1}, \boldsymbol{y}_{\mathrm{G}}^{u1,u2}, \boldsymbol{y}_{\mathrm{G}}^{u2,u2}, \boldsymbol{y}_{\mathrm{B}}^{u1,u1}, \boldsymbol{y}_{\mathrm{B}}^{u2,u1}, \boldsymbol{y}_{\mathrm{B}}^{u1,u2}, \boldsymbol{y}_{\mathrm{B}}^{u2,u2}, \boldsymbol{y}_{\mathrm{I}}^{u1,u1}, \boldsymbol{y}_{\mathrm{I}}^{u2,u1}, \boldsymbol{y}_{\mathrm{I}}^{u1,u2}, \boldsymbol{y}_{\mathrm{I}}^{u2,u2}\Big]^{\mathrm{T}}, \quad (23)$$

where $u1, u2$ index for the matrix position in the 2×2 CFA sub-group. The analytical optimum is

$$\boldsymbol{y}^{\mathrm{u}}{}_{\mathrm{opt}} = \left[\left(\boldsymbol{S}^{\mathrm{u}}\boldsymbol{C}^{\mathrm{u}}\right)^{\mathrm{T}} \left(\boldsymbol{S}^{\mathrm{u}}\boldsymbol{C}^{\mathrm{u}}\right) + \rho \right]^{-1} \left(\left(\boldsymbol{S}^{\mathrm{u}}\boldsymbol{C}^{\mathrm{u}}\right)^{\mathrm{T}} \boldsymbol{b}_1^{\mathrm{u}} + \rho \boldsymbol{b}_2^{\mathrm{u}} \right). \quad (24)$$

The optimum of $\boldsymbol{v}$ in line 5 is solved using shrinkage, as in [31]. Because $\boldsymbol{K}$, $\boldsymbol{D}$ and ∇ are all convolutional operations and therefore can be implemented efficiently in the frequency domain, we solve for $\boldsymbol{z}$ in the frequency domain [31]. Note also, in order to get fast convergence, we update ρ in each iteration [43]. In the reconstruction algorithm proposed above, the crosstalk matrix $\boldsymbol{C}$, the convolutional kernel $\boldsymbol{K}_i$, and the noise weighting coefficient $\boldsymbol{W}$ are assumed to be known. In the following, we describe how to pre-calibrate these parameters before reconstruction.

E. Calibration

We precalibrate the 4×4 crosstalk matrix $\boldsymbol{C}$ by solving a bi-convex optimization problem, alternately estimating the crosstalk matrix $\boldsymbol{C}$ and the ideal spectral response $\boldsymbol{Q}$ from the camera's spectral response $\boldsymbol{R}$.

In particular, we formulate the problem as

$$\underset{\boldsymbol{C},\boldsymbol{Q}}{\operatorname{argmin}} \ \| \boldsymbol{R} - \boldsymbol{C}\boldsymbol{Q} \|_2^2 + \alpha_1 \left\| \boldsymbol{I}_p \odot \boldsymbol{Q} \right\|_2^2 + \alpha_2 || \nabla \boldsymbol{Q} ||_2 \quad \text{subject to } \boldsymbol{C} \geq 0, \quad \boldsymbol{Q} \geq 0. \quad (25)$$

The first term is the least square fitting term, and the second term enforces the independence of the ideal spectral response of the R, G, B, I channels. The third term enforces the estimated ideal spectrum distribution $\boldsymbol{Q}$ to be smooth. Specifically, $\boldsymbol{R}$ is the column concatenated matrix of $[\boldsymbol{r}_{\mathrm{R}}, \boldsymbol{r}_{\mathrm{G}}, \boldsymbol{r}_{\mathrm{B}}, \boldsymbol{r}_{\mathrm{I}}]$, the spectral response matrix of the camera, and $\boldsymbol{Q}$ is the stacked matrix $[\boldsymbol{q}_{\mathrm{R}}, \boldsymbol{q}_{\mathrm{G}}, \boldsymbol{q}_{\mathrm{B}}, \boldsymbol{q}_{\mathrm{I}}]$, that is ideal spectral response matrix. $\boldsymbol{I}_p$ is a penalty matrix penalizing the spectral responses of $\boldsymbol{q}_{\mathrm{R}}, \boldsymbol{q}_{\mathrm{G}}, \boldsymbol{q}_{\mathrm{B}}$ outside the visible wavelength range and $\boldsymbol{q}_{\mathrm{I}}$ outside of NIR wavelength range. α_1, α_2 are the balancing coefficients for the second and third term, respectively.

$\boldsymbol{I}_p(a, b)$ is a penalty matrix, where a represents an index for the wavelength, and b is an index for the color channel. In the proposed method, we set $\boldsymbol{I}_p(a, b)$ for the R, G, B channel ($b = 1, 2, 3$) as

$$\boldsymbol{I}_p(a,b) = \begin{cases} 0 & \text{if } a \text{ in wavelength range } 300-700 \text{ nm} \\ 1 & \text{if } a \text{ in wavelength range } 700-900 \text{ nm}, \end{cases} \quad (26)$$

and for the NIR channel ($b = 4$)

$$\boldsymbol{I}_p(a,b) = \begin{cases} 1 & \text{if } a \text{ in wavelength range } 300-700 \text{ nm} \\ 0 & \text{if } a \text{ in wavelength range } 700-900 \text{ nm.} \end{cases} \quad (27)$$

∇ denotes the derivative operator along the wavelength dimension, i.e. the first dimension of matrix $\boldsymbol{Q}$. The problem (25) is a biconvex problem and can be solved by alternatively optimizing $\boldsymbol{C}$ and $\boldsymbol{Q}$, while fixing the respective other. Each subproblem is a convex optimization problem and is solved using a first-order method.

In our implementation, we multiply the crosstalk matrix $\boldsymbol{C}$ with a white balance diagonal matrix so that the reconstructed latent image is white balanced. The white-balance coefficients of the RGB channels are calibrated by capturing a uniform scene, e.g. white wall, and estimating the mean ratio of the RGB channels. The white balance coefficient of the NIR channel is set to 1.

To estimate $\boldsymbol{K}$, we use the tiling calibration pattern proposed in [44] and estimate the in-focus plane blur of each channel with the estimation method in [45]. As the most severe crosstalk occurred from NIR to all three RGB channels, we use RGB-cutoff and IR-cutoff filters for the PSF estimation captures. We capture the target image with an NIR-cutoff filter to calibrate $\boldsymbol{K}_{\mathrm{R}}, \boldsymbol{K}_{\mathrm{G}}, \boldsymbol{K}_{\mathrm{B}}$, and using an RGB-cutoff filter to calibrate $\boldsymbol{K}_{\mathrm{I}}$. In Fig. 1(i), we show the estimated blur kernel of different channels estimated from the center region of the image. As shown, the blur of the NIR channel differs substantially from the blur in the RGB channels.

We estimated the noise level of each channel w_i with the method in [46], and the corresponding weighting matrix is

$$\boldsymbol{W}[t,t] = w_i, \text{if pixel at } t \text{ has color filter } i,\ i \in \{\mathrm{R},\mathrm{G},\mathrm{B},\mathrm{I}\}. \quad (28)$$

Having described the proposed reconstruction method, next we validate our method using both numerical simulations and physical measurements.

IV. Assessment

This section evaluates the proposed approach on synthetic and measured experimental data. Comparisons to existing methods validate that convolutional sparse RGB+NIR reconstruction outperforms the state-of-the-art quantitatively and qualitatively in a wide range of scenarios, including challenging structured illumination cases.

A. Synthetic Evaluation

1) Data Generation and Calibration: We first evaluate the proposed method in simulation. To this end, we synthesize measurements $\boldsymbol{j}$ from the hyperspectral database [47]. From a precalibrated camera response function $r(\lambda)$ of a representative reference camera we estimate the crosstalk matrix $\boldsymbol{C}$ by solving the bi-convex problem in Eq. (25), with $\alpha_1 = 0.5$ and $\alpha_2 = 0.1$. The convergence plot of the objective function along the iteration number is shown in Fig. 1(c); the optimization converges in approximately 15 iterations. The estimated crosstalk matrix is shown in Fig. 1(h) which is nearly diagonal except for the fourth column, indicating the influence of the NIR channel to the RGB channels. The blur kernels $\boldsymbol{K}_{\mathrm{R}}, \boldsymbol{K}_{\mathrm{G}}, \boldsymbol{K}_{\mathrm{B}}, \boldsymbol{K}_{\mathrm{I}}$ of each channel are calibrated from chart images, as described in Sec.III-E. Note that for the scenes with structured NIR illumination, we replaced the NIR channel from the hyperspectral input data with synthetically generated high-frequency dot patterns that accurately resemble the ones we measured from the Intel RealSense measurements.

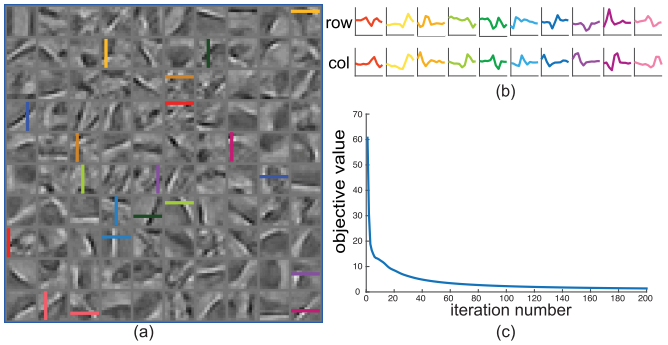


Fig. 2. (a) Learned convolutional kernels. (b) Horizontal and vertical scanlines picked from the dictionaries in (a). (c) Convergence plot of the learning procedure, showing a monotonically-decreasing decay.

With the simulated measurements at hand, we next learn a dictionary of convolutional kernels from the "fruit" dataset of [34]. The learned convolutional codes are shown in Fig. 2(a). Each of the 100 kernels is of size 11×11. The training error is plotted in Fig. 2(c), which shows that the learned filters are converged. Some of the learned dictionary elements represent low-frequency structure in the image, while others model high-frequency structure, such as edges or corners. Figs. 2(a, b) highlight filters aid the reconstruction of high-frequency content in the proposed convolutional sparse RGB+NIR imaging method.

2) Qualitative Evaluation: Given the measurements and learned priors, we compare the proposed method against three existing reconstruction methods. As a baseline we compare against naive bi-cubic upsampling. Furthermore, we also compare the proposed method against two state-of-the-art reconstruction approaches: a hand-crafted recovery algorithm [21] and an optimization-based MAP estimation method [6]. Figs. 3 and 4 show qualitative results for NIR ambient light and high-frequency structured NIR illumination. In all scenarios, the proposed approach achieved substantially improved image quality in all spectral channels compared to the reference methods. In particular, high-frequency color structures around edges and small features are accurately recovered. In the structured illumination scenarios, severe artefacts are observable in the RGB channels of all existing approaches. The proposed method is the only one that achieves high quality in this challenging illumination case.

3) Parameter Evaluation: Next, we analyze the effect of different parameters of the proposed reconstruction algorithm. We performed a parameter sweep of β and γ, and plot the best recovery performance with respect to β and γ, shown in Figs. 5(a-d), and at different noise levels, corresponding to a range of 0.006, 0.013, 0.019, 0.042, 0.06 in standard deviation.

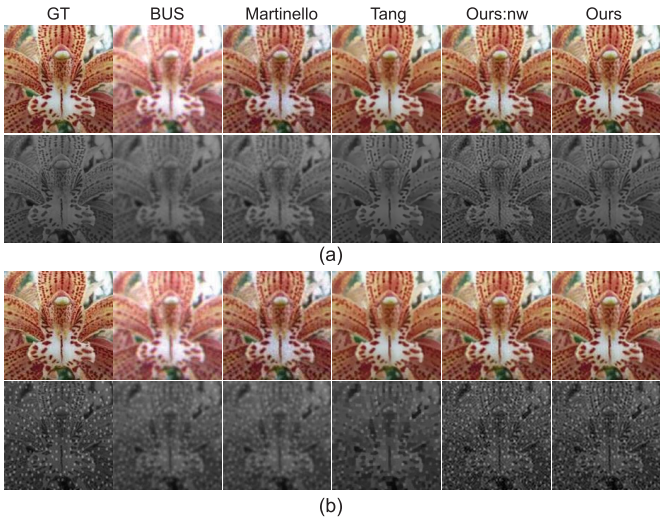


Fig. 3. Recovery results for the "flowers" scene with different methods, without structured NIR illumination (a) and with structured illumination (b). GT, BUS, Martinello, Tang, Ours:nw and Ours denote the ground truth and the reconstructed results of basic bilinear upsampling method, [21], [6], the proposed method without and with weighting $\boldsymbol{W}$.

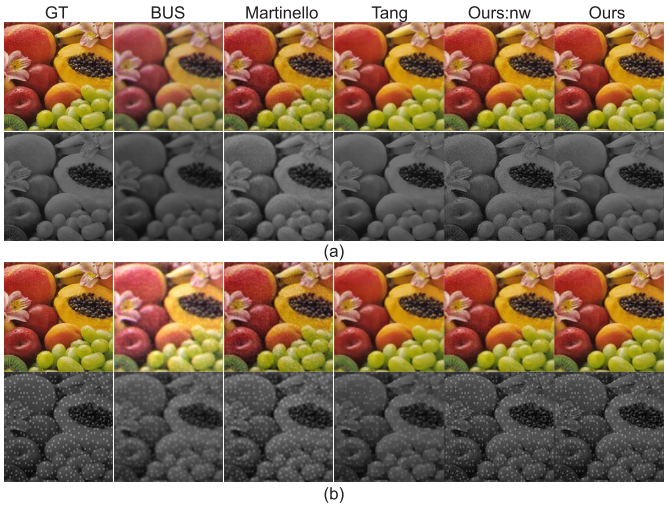


Fig. 4. Recovery results for the "fruits" scene. The same methods as in Fig. 3 are compared.

As the noise level increases, the best performance parameters β_{best} and γ_{best}, marked with a red diamond, increase correspondingly to balance the three terms in the optimization problem. As the ground truth image does not change, the ratio of β_{best} and γ_{best} are approximately constant, which translates the robustness of the parameter selection when estimating the noise level ahead of the RGB+NIR recovery. In other words, the main purpose of β and γ is to balance the second prior term and the third prior term. To demonstrate that this behavior generalizes across different illumination scenarios, we performed the same analysis with structured NIR illumination (Figs. 5(a, b)) and with structured NIR illumination (Figs. 5(c, d)).

Furthermore, comparing Figs. 5(a) and (c) with Figs. 5(b) and (d), we further observe that the prior coefficient β plays a key role in the reconstruction. When β is

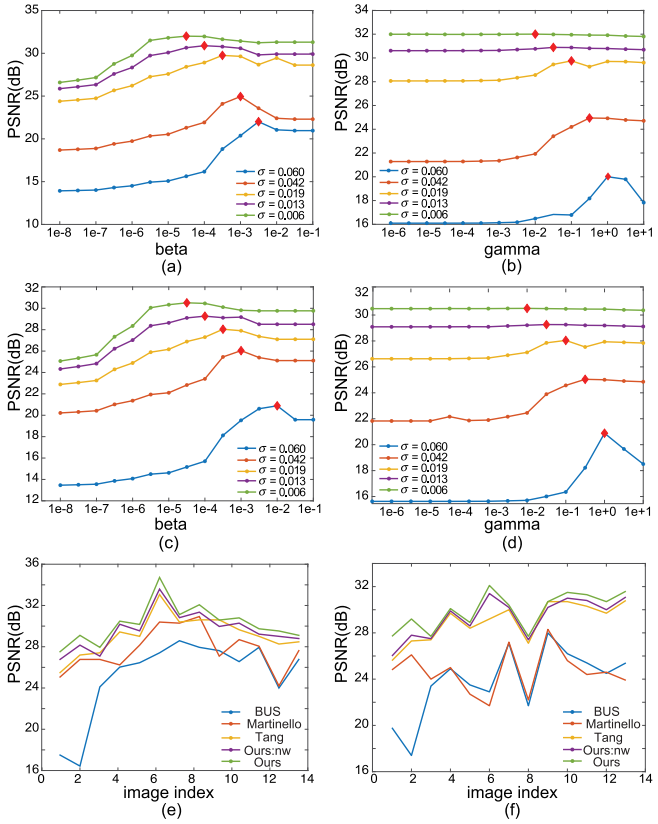


Fig. 5. Parameters analysis. (a) Reconstruction performance with respect to different β, with no structured NIR illumination. (b) Reconstruction quality for different γ, with no structured NIR illumination. (c) and (d) simulate the same setting as (a) and (b) but with structured NIR illumination. (e, f) Reconstruction quality for different methods using the same method shortcuts as in Fig. 3.

not in a proper range, quality is poor no matter what the parameter γ is set to. Intuitively, the parameter γ fine-tunes the reconstruction performance when β is in the proper range.

In addition to the parameter experiments discussed above, we also explored the effect of the noise weighting. Noise weighting balances the objective term and it is used for whitening the objective of the four channels in the first data fidelity term in Eq. (15). For a fixed noise level of standard deviation 0.019, we choose the deconvolution parameters of the method [21] to be $\alpha = 2e3$ and $\lambda = 2/3$ (deconvolution method in [5]), $\sigma = 1/400$, $\tau = 40$, and $\alpha = 5$ for method [6] and $\beta = 5e-3$, $\gamma = 1e-1$ for the proposed method. Note that these parameters are the best parameters of an extensive parameter sweep for the given image. As shown in Fig. 3 and Fig. 4, without the noise level weighting, the reconstruction performance of the proposed method in the NIR and RGB channels are unbalanced and the NIR channel reconstruction is noisy. Figs. 5(e) and (f) show these comparisons for different hyperspectral input images. With noise weighting, the described approach performs the best. While without noise weighting, the proposed method sometimes is only on par with existing methods, highlighting that this weighting is essential for balancing convolutional sparse coding across channels.

TABLE II
PSNR COMPARISONS (dB)

Illumination	Without Structured Illumination				With Structured Illumination			
Methods	BUS	Martinello et. al	Tang et. al	Proposed method	BUS	Martinello et. al	Tang et. al	Proposed method
Flowers	18.0	25.1	25.5	**27.4**	19.8	24.8	25.6	**27.7**
Fruits	17.0	26.7	27.1	**28.9**	17.4	26.1	27.3	**29.2**
Books	24.2	26.7	27.3	**27.8**	23.4	24.0	27.4	**27.7**
EsserBlocks	26.0	26.2	29.2	**30.2**	24.9	25.0	29.7	**30.1**
EsserCalib	26.4	28.0	28.8	**29.9**	23.5	22.7	28.4	**28.9**
FreshFruitsBasket	27.3	30.1	32.6	**34.2**	22.9	21.7	29.2	**32.1**
FruitPlatter	28.4	30.0	30.1	**30.8**	27.1	27.2	30.0	**30.4**
HiResFemale2	27.8	30.6	30.3	**31.7**	21.7	22.2	27.1	**27.7**
MCCfullresWithNIR	27.5	27.0	**30.3**	**30.3**	28.0	28.3	30.7	**30.7**
SanFrancisco	26.5	28.5	29.4	**30.5**	26.2	25.6	30.7	**31.5**
StanfordDish	27.8	27.9	28.8	**29.5**	25.4	24.4	30.3	**31.3**
StanfordMemorial	24.1	24.3	28.1	**29.3**	24.5	24.6	29.7	**30.7**
StanfordTower	26.7	27.5	28.3	**28.9**	25.4	23.9	30.8	**31.6**

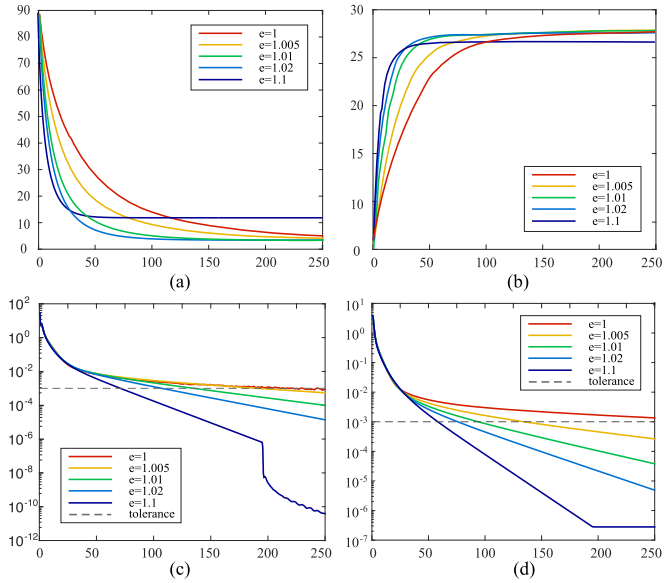


Fig. 6. Analysis of the step size e in Algorithm 1. We process the flower scene from Fig. 3(a) here and show: (a) the objective values, (b) PSNR, and the ℓ_2-norm of the primal (c) and dual (d) residuals [42].

We also analyzed the effects of the step size parameter e in Algorithm. 1. As discussed in [48], dynamically adapting the penalty parameter ρ can substantially improve convergence in practice. Fig. 6 compares convergence behavior for varying e, with all parameters except for e fixed. For $e > 1$, the ℓ_2 norm of the primal and the dual residuals [42] are reduced at a higher rate with increasing $e > 1$. Specifically, while for $e = 1$, about 250 iterations are required for both the primal residuals and dual residuals to converge below $1e-3$, using $1 < e < 1.1$ required only half of the iterations to converge to the same accuracy. Therefore, updating the penalty parameter ρ using $e > 1$ substantially improves convergence. However, when e becomes large, the ρ update rate is very fast, hence consensus is enforced aggressively and the optimization algorithm is seemingly stuck, making only very slow progress. For the experimental results, we choose a small $e = 1.01$ to speed up the convergence of ADMM.

Finally, we provide recipes for selecting the remaining objective hyper-parameters. To select β and γ we run the proposed method on separate test datasets. First, we optimize β in the large range $[10^{-6}, 1]$ with fixed $\gamma = 10^{-3}$. Subsequently, we optimize for γ in the range $[10^{-3}, 1]$ with β now fixed. The optimal size of the dictionary kernel depends on the scale of the images. Adopting the setting from [31], we use filters of spatial extent 11×11.

4) Quantitative Evaluation on Benchmark Datasets: Performance results of all methods on a benchmark dataset are shown in Table II and Fig. 7. When there is no structured NIR illumination, the proposed method performs much better in preserving details, such as the text in the book scene, the wax crayon, the toy eye, the banana textures, the small holes on the strawberry, the eyebrows in the human faces, the outdoor tree textures, the Stanford dish, and the texture on the Church. When the image is piece-wise smooth, the described method performs similarly to [6] as their TV prior is sufficient in these cases. Furthermore, we notice that [21] introduces structured illumination artefacts into the RGB channel, severely affecting the reconstruction quality. In contrast, for structured NIR illumination, the proposed method preserves the high-frequency details introduced by the NIR illumination. The NIR reconstruction of the proposed method is the most accurate compared to existing previous approaches. Having validated the RGB+NIR reconstruction approach using numerical simulations, we next assess the proposed method using experimental measurements.

B. Physical Experiments

Reconstruction results from captured measurements without structured illumination are shown in Figs. 8, 9, and 10. For the example shown in Fig. 8 we compare the proposed method with bilinear upsampling, the method of Martinello *et al.* [21], and Tang *et al.* [6]. The bilinear upsampled results suffer from severe color artefacts because an appropriate model for the color crosstalk is missing. We have included magnified crops on the right of Fig. 8(d), demonstrating the details we recover are not noise but the actual texture in the bill. The proposed method is capable of reconstructing fine scene

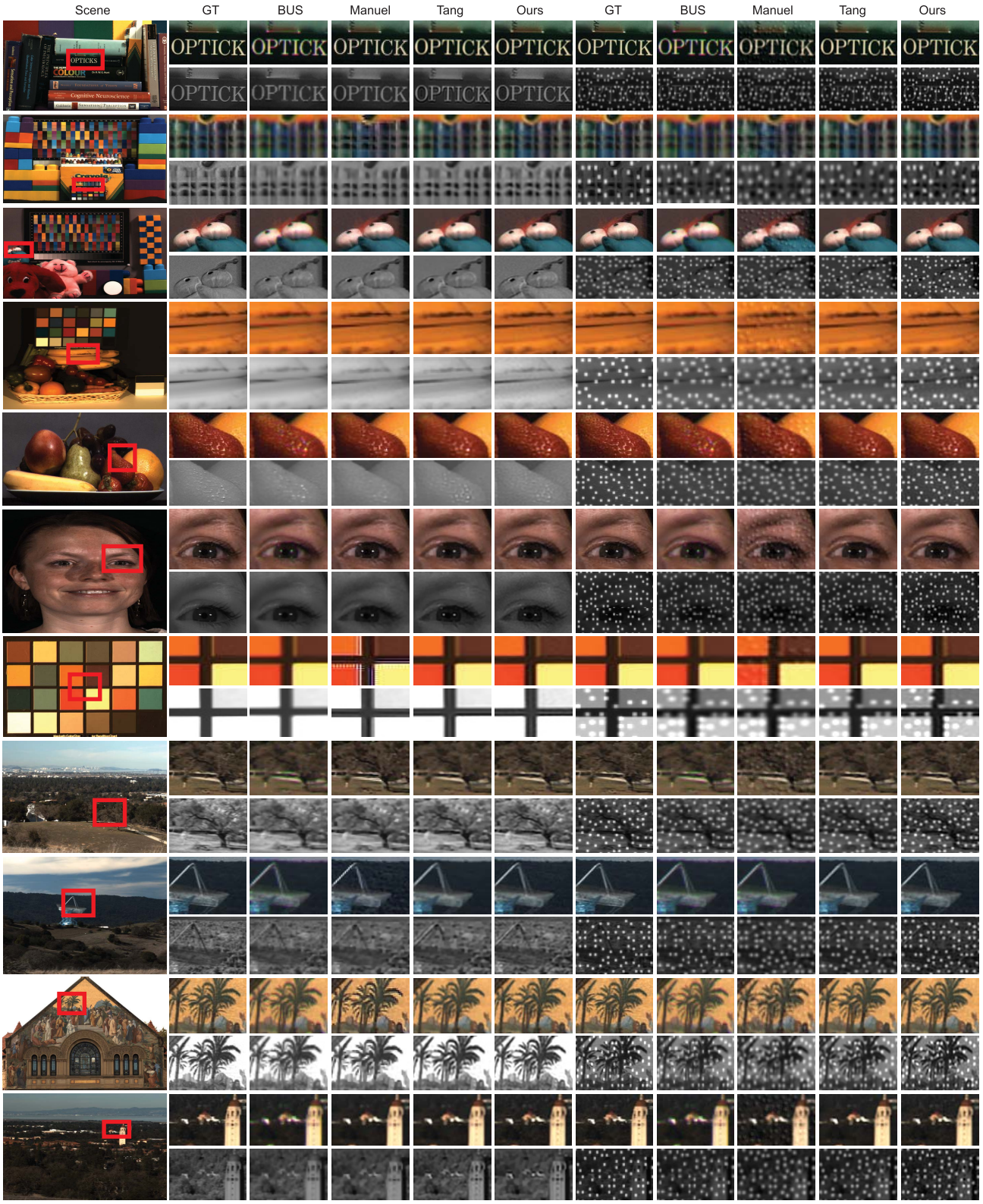


Fig. 7. Reconstruction results on the hyperspectral database described in the text, without (columns 2-6) and with (columns 7-11) structured NIR illumination.

structures of the bills and table that are not recovered by Tang's or Martinello's approaches. Specifically, [6] tends to smooth out the IR channel when there is small intensity contrast, whereas the proposed CSC method is able to recover sharper edges of the number 5 on the bill and the scratches on the table. The approach of [21] tends to recover noisy RGB and IR channels, where the CSC method was able to recover high-quality RGB and NIR channel.

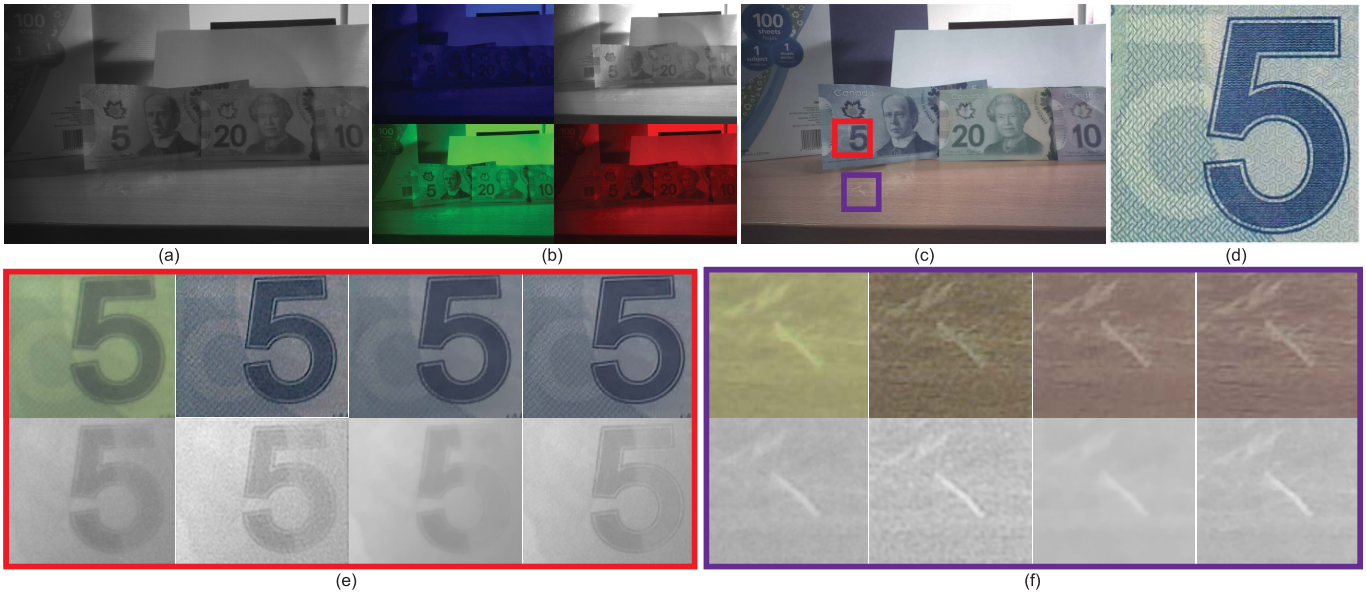


Fig. 8. Indoor scene without structured illumination. (a) Captured raw image, (b) extracted color channels, (c) the reconstructed RGB image of the proposed method. (d) Cropped photography of the figure 5 on the bill. We compare reconstructions with bilinear upsampling, Martinello *et al*., Tang *et al*. and the proposed approach. In each close-up in (e) and (f), the figures we show are bilinear upsampling (first column), Martinello et al. (second column), Tang *et al*. (third column), and the proposed method (fourth column) of RGB (first row) and NIR (second row) channel respectively.

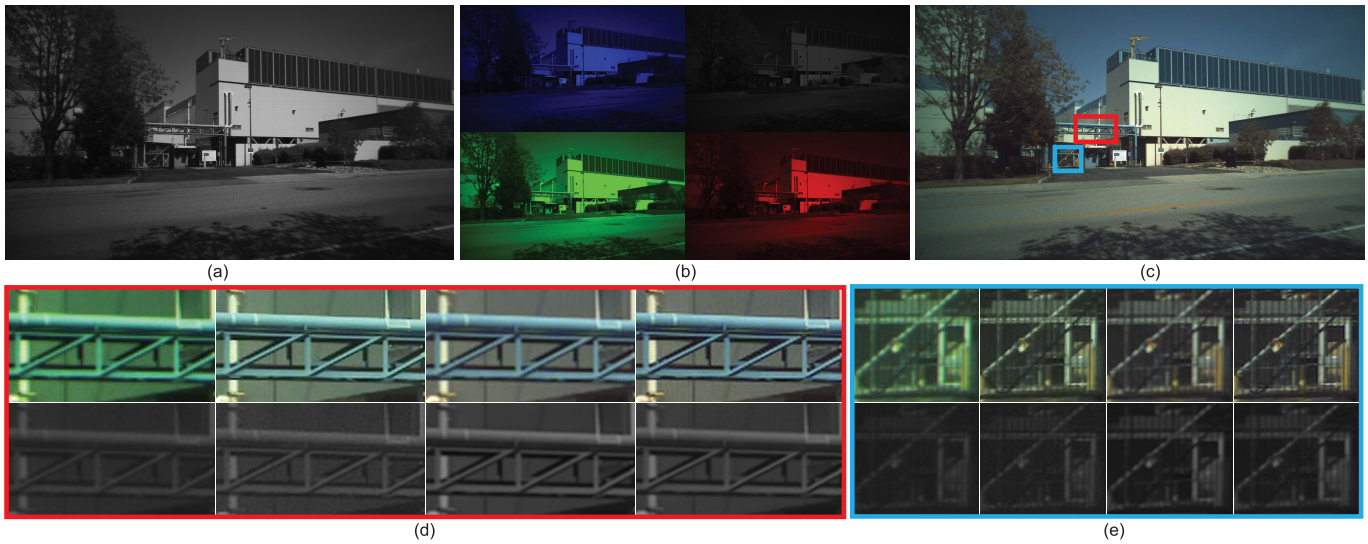


Fig. 9. Outdoor scene without structured illumination. The denotation of (a)-(e) is the same as Fig. 8 (a)-(c) and (e)-(f).

In Figs. 9 and 10, we show several results captured with an OmniVision OV4682 sensor. As shown in the magnified crops for the indoor scene from Figs. 10(d)-(f), fine detail, such as the human eye region, the stairway, and the text on the newspaper are recovered in the RGB and NIR channels with sharper edges and more accurate colors compared to existing approaches. Fig. 9 shows results from an outdoor scene with natural sunlight illumination. In contrast to the other methods, the proposed approach is capable of recovering sharp image features and accurate colors of both the crossbeam (Fig. 9(e)), and the iron fence and the scene behind it in Fig. 9(f). All these experiments without structured NIR illumination have been computed with the following algorithm parameters: $\beta = 0.001$, $\gamma = 0.01$, and $\rho_0 = 1$.

We have also evaluated the robustness of the proposed reconstruction method in relation to structured illumination in the NIR domain. In Fig. 11 the scene is illuminated with ambient light but also with a high-frequency NIR dot pattern. This setup emulates a condition where scene depth can be recovered from the NIR channel along with a high-quality RGB image. The result from bilinear upsampling, shown in Fig. 11 does not remove the high-frequency NIR pattern in the color channels, leading to severe chromatic artefacts, in contrast to the proposed method, which recovers high-quality RGB reconstruction results. As shown in Figs. 11(e)-(f), the structured illumination contaminates the RGB channels, which can be seen in the bilinear upsampled result, for example the speckle-like pattern on the toy's back and the checkerboard.

Fig. 10. Indoor scene without structured NIR illumination. (a) Captured raw image of an indoor scene, (b) extracted color channels, (c) the reconstructed RGB image of the proposed method. We compare reconstructions of bilinear upsampling, Martinello *et al*. (second column), Tang *et al*. (third column), and the proposed method (fourth column) of RGB (first row) and NIR (second row) channel respectively. The smaller close-up are shown on the right of each figure for further comparison.

In contrast to Tang *et al.* [6], the proposed approach recovers more details, such as the water ripples on the oil paints and the leaves. Closely following the trend from the simulation results, the details in the NIR channel are smoothed out by Tang's method. The method of Martinello *et al.* [21] cannot handle the case where there is structured illumination. These results are encouraging and could inform future approaches to joint RGB and depth imaging from structured NIR illumination. For the above experiments, we use the following algorithmic parameters: $\beta = 0.0001$, $\gamma = 0.1$, and $\rho_0 = 1$. For the physical experiments, we use the noise weighting coefficient $[w_R, w_G, w_B, w_I] = [1, 1, 1, 0.5]$.

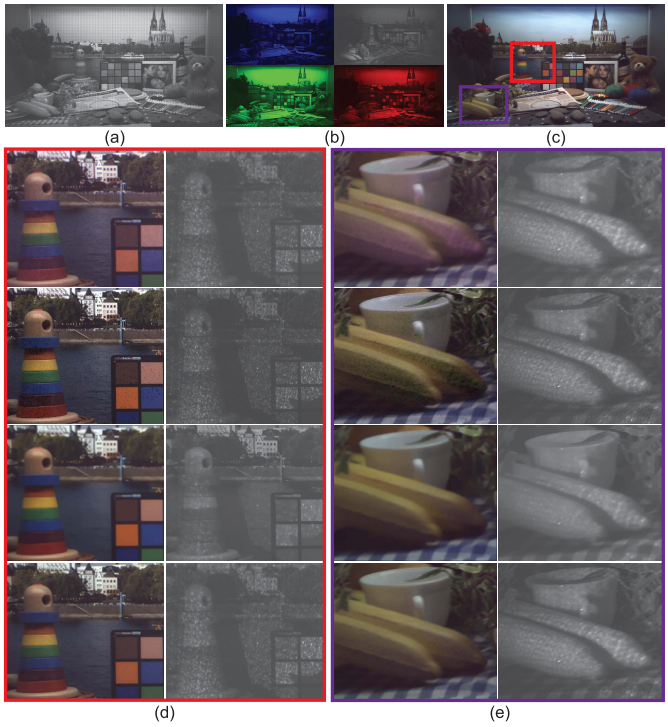


Fig. 11. Indoor scene with structured NIR illumination. The denotation of (a)-(e) is the same as in Fig. 8, except that in each close-up, the columns correspond to color channels, first column for RGB and second column for NIR. Four rows correspond to: bilinear upsampling (first row), Martinello *et al*. (second row), Tang *et al*. (third row) and the proposed method (fourth row).

We would like to point out that for a given noise level, we have fixed the parameters for the complete dataset. The relative weighting between the data term and prior depends on the standard deviation of the noise and therefore has to be tuned differently for different noise scenarios. Note again that this is not specific to the proposed method, but applies to all comparable MAP estimation methods. We adopt a computationally cheap noise estimation method [49] to estimate the noise parameters.

Having validated the proposed image formation model and reconstruction method in this section, we next discuss the benefits and limitations of the proposed approach, and identify potential future directions building on this work.

V. Discussion

In summary, we have presented a novel framework for recovering multi-channel image data from spatially multiplexed measurements. The learned convolutional sparse prior for RGB+NIR images, represents a stronger prior than gradient priors, which are often used in image reconstruction and are suggested in [6]. Note that rather than using just two hard-coded, engineered $[2 \times 1]$ derivative filters, we learn 100 filters with $[11 \times 11]$ coefficients. It becomes obvious that our model is orders of magnitude more expressive than previous approaches. We have evaluated the proposed method in simulation and with data captured by a RGB+NIR sensor. We demonstrate improved reconstruction quality compared to state-of-the-art approaches. Furthermore, we have evaluated camera systems that use structured illumination in the NIR range. Such systems are increasingly used by commercial depth cameras that scan the scene using invisible patterns. The structured illumination patterns are observed as artefacts in the other channels, making it difficult to recover RGB images alongside the coded NIR channel. We demonstrate high-quality RGB image recovery for these challenging scenarios.

A. Limitations

The resolution of all captured images is 1520×2688 and we split them up into 4×8 blocks, each 380×336 in size. We subsequently run our reconstruction on each block for 50 iterations. On an Intel Xeon E5 machine, our unoptimized MATLAB implementation runs currently in about 10 minutes for each block. All individual blocks are run in parallel. We also compared our approach to a conventional TV prior-based reconstruction, but without the denoising and inpainting strategies described in [6]. This approach took about 25 iterations to converge in 8.5 minutes per block. The proposed method shares with other optimization-based methods that it is not computationally cheap. While an immediate solution is cloud processing, a more practical mobile approach are emerging low-power image processing units, such as the Movidius Myriad 2 which evaluates the VGG16 convolutional net at real-time frame rates at low power consumption.[3] Note that the proposed method shares striking similarities with convolutional nets, which makes it map well to such emerging architectures.

B. Future Work

The proposed image formation model does not model saturation, a limitation shared with [6]. Saturated data can be masked in the objective function, which then will be effectively be inpainted by the masked reconstruction approach. To make the proposed framework practical for widespread mobile and robotic applications, an implementation using field-programmable gate arrays or an application-specific integrated circuit is left for future work. In the near future, we are planning to incorporate a depth-from-structured-NIR-illumination reconstruction step into our framework. This is possible by either accessing calibration data on existing consumer devices or by building a structured illumination projector–camera system from scratch. Although interesting, we leave this engineering effort for future work. Finally, we believe that exploring hierarchical filtering would also be a very interesting direction of future research.

Acknowledgment

The authors would like to thank Intel Corporation for providing the raw RGB+NIR data and H. Tang for providing comparison results on synthetic and measured data.

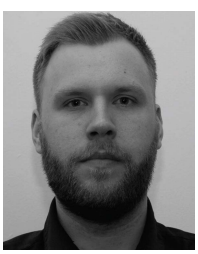
References

[1] M. Vollmer and K.-P. Möllmann, *Infrared Thermal Imaging: Fundamentals, Research and Applications*. Weinheim, Germany: Wiley, 2010.

[3] http://www.movidius.com/solutions/vision-processing-unit

[2] M. Brown and S. Süsstrunk, "Multi-spectral SIFT for scene category recognition," in *Proc. IEEE CVPR*, Jun. 2011, pp. 177–184.
[3] K. Tanaka, Y. Mukaigawa, Y. Matsushita, and Y. Yagi, "Descattering of transmissive observation using parallel high-frequency illumination," in *Proc. IEEE ICCP*, Apr. 2013, pp. 1–8.
[4] L. Schaul, C. Fredembach, and S. Süsstrunk, "Color image dehazing using the near-infrared," in *Proc. IEEE ICIP*, Nov. 2009, pp. 1609–1612.
[5] D. Krishnan and R. Fergus, "Dark flash photography," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–96, 2009.
[6] H. Tang, X. Zhang, S. Zhuo, F. Chen, K. N. Kutulakos, and L. Shen, "High resolution photography with an RGB-infrared camera," in *Proc. IEEE ICCP*, Apr. 2015, pp. 1–10.
[7] Y. M. Lu, C. Fredembach, M. Vetterli, and S. Süsstrunk, "Designing color filter arrays for the joint capture of visible and near-infrared images," in *Proc. IEEE ICIP*, Nov. 2009, pp. 3797–3800.
[8] G. Langfelder, T. Malzbender, A. F. Longoni, and F. Zaraga, "A device and an algorithm for the separation of visible and near infrared signals in a monolithic silicon sensor," *Proc. SPIE*, vol. 7882, p. 788207, Jan. 2011.
[9] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Simultaneous capturing of RGB and additional band images using hybrid color filter array," *Proc. SPIE*, vol. 9023, p. 90230V, Mar. 2014.
[10] 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," in *Proc. IEEE ICASSP*, May 2013, pp. 1646–1650.
[11] D. Fofi, T. Sliwa, and Y. Voisin, "A comparative survey on invisible structured light," *Proc. SPIE*, vol. 5303, pp. 90–98, May 2004.
[12] J. Garcia and Z. Zalevsky, "Range mapping using speckle decorrelation," U.S. Patent 7 433 024, Oct. 7, 2008.
[13] A. Shpunt and Z. Zalevsky, "Depth-varying light fields for three dimensional sensing," U.S. Patent 8 050 461, Nov. 1, 2011.
[14] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, Feb. 2012.
[15] J. Smisek, M. Jancosek, and T. Pajdla, "3D with Kinect," in *Consumer Depth Cameras for Computer Vision*. London, U.K.: Springer, 2013, pp. 3–25.
[16] R. Kimmel, "Demosaicing: Image reconstruction from color CCD samples," *IEEE Trans. Image Process.*, vol. 8, no. 9, pp. 1221–1228, Sep. 1999.
[17] X. Li, B. Gunturk, and L. Zhang, "Image demosaicing: A systematic survey," *Proc. SPIE*, vol. 6822, p. 68221J, Jan. 2008.
[18] O. Losson, L. Macaire, and Y. Yang, "Comparison of color demosaicing methods," *Adv. Imag. Electron Phys.*, vol. 162, pp. 173–265, Oct. 2010.
[19] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.
[20] B. Sajadi, A. Majumder, K. Hiwada, A. Maki, and R. Raskar, "Switchable primaries using shiftable layers of color filter arrays," *ACM Trans. Graph.*, vol. 30, no. 4, p. 65, 2011.
[21] M. Martinello *et al.*, "Dual aperture photography: Image and depth from a mobile camera," in *Proc. IEEE ICCP*, Apr. 2015, pp. 1–10.
[22] Q. Tian, S. Lansel, J. E. Farrell, and B. A. Wandell, "Automating the design of image processing pipelines for novel color filter arrays: Local, linear, learned (L^3) method," *Proc. SPIE*, vol. 9023, p. 90230K, Mar. 2014.
[23] B. A. Olshausen and D. J. Field, "Natural image statistics and efficient coding," *Netw., Comput. Neural Syst.*, vol. 7, no. 2, pp. 333–339, 1996.
[24] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.
[25] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proc. NIPs*, 2006, pp. 801–808.
[26] M. Aharon, "Overcomplete dictionaries for sparse representation of signals," Ph.D. dissertation, Dept. Comput. Sci., Technion—Israel Inst. Technol., Haifa, Israel, 2006.
[27] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59–73, Jul. 2011.
[28] X. Lin, G. Wetzstein, Y. Liu, and Q. Dai, "Dual-coded compressive hyperspectral imaging," *Opt. Lett.*, vol. 39, no. 7, pp. 2044–2047, 2014.
[29] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 46:1–46:12, 2013.
[30] H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," in *Proc. IEEE CVPR*, Jun. 2013, pp. 391–398.
[31] F. Heide, W. Heidrich, and G. Wetzstein, "Fast and flexible convolutional sparse coding," in *Proc. IEEE CVPR*, Jun. 2015, pp. 5135–5143.
[32] B. Choudhury, R. Swanson, F. Heide, G. Wetzstein, and W. Heidrich, "Consensus convolutional sparse coding," in *Proc. IEEE ICCV*, Apr. 2017, pp. 1–8.
[33] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.
[34] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE CVPR*, Jun. 2010, pp. 2528–2535.
[35] K. Kavukcuoglu, P. Sermanet, Y.-L. Boureau, K. Gregor, M. Mathieu, and Y. LeCun, "Learning convolutional feature hierarchies for visual recognition," in *Proc. NIPs*, 2010, pp. 1090–1098.
[36] C. Osendorfer, H. Soyer, and P. van der Smagt, "Image super-resolution with fast approximate convolutional sparse coding," in *Proc. NIPs*, 2014, pp. 250–257.
[37] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," in *Proc. IEEE CVPR*, Dec. 2015, pp. 1823–1831.
[38] Y. Zhou, H. Chang, K. Barner, P. Spellman, and B. Parvin, "Classification of histology sections via multispectral convolutional sparse coding," in *Proc. IEEE CVPR*, Jun. 2014, pp. 3081–3088.
[39] X. Hu *et al.*, "Robust and accurate transient light transport decomposition via convolutional sparse coding," *Opt. Lett.*, vol. 39, no. 11, pp. 3177–3180, 2014.
[40] F. Heide, L. Xiao, A. Kolb, M. B. Hullin, and W. Heidrich, "Imaging in scattering media using correlation image sensors and sparse convolutional coding," *Opt. Exp.*, vol. 22, no. 21, pp. 26338–26350, 2014.
[41] A. Serrano, F. Heide, D. Gutierrez, G. Wetzstein, and B. Masia, "Convolutional sparse coding for high dynamic range imaging," *Comput. Graph. Forum*, vol. 35, no. 2, pp. 153–163, May 2016.
[42] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[43] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 644–658, Mar. 2015.
[44] N. Joshi, R. Szeliski, and D. J. Kriegman, "PSF estimation using sharp edge prediction," in *Proc. IEEE CVPR*, Jun. 2008, pp. 1–8.
[45] T. Michaeli and M. Irani, "Blind deblurring using internal patch recurrence," in *Proc. IEEE ECCV*, 2014, pp. 783–798.
[46] X. Liu, M. Tanaka, and M. Okutomi, "Noise level estimation using weak textured patches of a single noisy image," in *Proc. IEEE ICIP*, Sep./Oct. 2012, pp. 665–668.
[47] T. Skauli and J. Farrell, "A collection of hyperspectral images for imaging systems research," *Proc. SPIE*, vol. 8660, p. 86600C, Feb. 2013.
[48] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. NIPs*, 2011, pp. 612–620.
[49] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian–Gaussian noise modeling and fitting for single-image raw-data," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1737–1754, Oct. 2008.

Xuemei Hu received the B.Sc. degree with the Department of Automation, Tsinghua University, China, in 2013, where she is currently pursuing the Ph.D. degree. She is also a Visiting Ph.D. Student with Stanford University. Her research interests are centered on computational photography.



Felix Heide received the B.Sc. and M.Sc. degree from the University of Siegen, and the Ph.D. degree from The University of British Columbia. He is currently a Post-Doctoral Scholar with Stanford University. His research interests lie in computational imaging, optimization, and displays.

Qionghai Dai received the B.Sc. degree in mathematics from Shanxi Normal University, China, in 1987, and the M.E. and Ph.D. degrees in computer science and automation from Northeastern University, China, in 1994 and 1996, respectively. Since 1997, he has been with the Faculty of Tsinghua University, Beijing, China, where he is currently a Professor and the Director of the Multi-dimension and Multi-scale Computational Photography Laboratory. His research areas include 3D video, computer vision, computational photography, and microscopy.

Gordon Wetzstein received the Diplom degree in media system science from the Bauhaus University at Weimar in 2006 and the Ph.D. degree in computer science from The University of British Columbia in 2011. He was a Research Scientist with the MIT Media Lab from 2011 to 2014. He has been an Assistant Professor with Stanford University since 2014. His research focuses on computational imaging and display.