High-resolution Hyperspectral Imaging via Matrix Factorization

Rei Kawakami*
The University of Tokyo
rei@cvl.iis.u-tokyo.ac.jp

John Wright
Microsoft Research Asia
jowrig@microsoft.com

Yu-Wing Tai
KAIST
yuwing@cs.kaist.ac.kr

Yasuyuki Matsushita
Microsoft Research Asia
yasumat@microsoft.com

Moshe Ben-Ezra
Microsoft Research Asia
mosheb@microsoft.com

Katsushi Ikeuchi
The University of Tokyo
ki@cvl.iis.u-tokyo.ac.jp

Abstract

Hyperspectral imaging is a promising tool for applications in geosensing, cultural heritage and beyond. However, compared to current RGB cameras, existing hyperspectral cameras are severely limited in spatial resolution. In this paper, we introduce a simple new technique for reconstructing a very high-resolution hyperspectral image from two readily obtained measurements: A lower-resolution hyperspectral image and a high-resolution RGB image. Our approach is divided into two stages: We first apply an unmixing algorithm to the hyperspectral input, to estimate a basis representing reflectance spectra. We then use this representation in conjunction with the RGB input to produce the desired result. Our approach to unmixing is motivated by the spatial sparsity of the hyperspectral input, and casts the unmixing problem as the search for a factorization of the input into a basis and a set of maximally sparse coefficients. Experiments show that this simple approach performs reasonably well on both simulations and real data examples.

1. Introduction

The spectrum of scene radiance is characterized by a distribution of intensity at each wavelength. Conventional RGB cameras perform a very gross quantization of this distribution, integrating the product of the intensity with three basis functions that are significantly spread over wavelength. Inevitably, information about the scene radiance, and hence about the physical constituents of the scene, is lost in this process.

The goal of hyperspectral imaging is to acquire much more faithful representations of the scene radiance. The typical output of a hyperspectral imaging system is a set of intensity values that represent the integrals of the radiance against many basis functions, each of which is well-localized in wavelength. The availability of such a detailed, physical representation of the scene is critical for some applications. For example, in medical image analysis, full information about the scene radiance can be used to detect anomalies that are not distinguishable from RGB images alone. Also, faithful color reproduction is critical for museums to record their artwork. Moreover, the use of hyperspectral imagery has been shown to enhance the performance of a number of computer vision tasks, including segmentation, tracking, and recognition [18].

Unfortunately, this utility comes at a price: the spatial resolution of current hyperspectral imaging systems is severely limited compared to that of RGB cameras. Recording images with a number of spectral bands requires not only a larger number of exposures but also longer exposure times to ensure good signal-to-noise ratio in low-light conditions. It is therefore very interesting to investigate new imaging mechanisms which can acquire detailed radiance information without sacrificing spatial resolution.

Contributions: In this paper, we develop a practical high-resolution hyperspectral imaging system based on a high-resolution RGB camera and low-resolution hyperspectral camera. We introduce a two-step process for combining these two measurements to estimate a high-resolution hyperspectral image. First, using a principled solution to the *spectral unmixing* problem [13], we obtain an optimal set of basis functions for representing the reflectances of the various materials in the scene. Motivated by the simple observation that there are likely to only be a few materials contributing to each pixel in the hyperspectral image, we cast the unmixing problem as one of sparse matrix factorization, and solve it via ℓ^1-minimization. We use the basis obtained from this method, together with the high-resolution RGB observation to reconstruct the spectrum at each location. The proposed method can efficiently obtain very accurate spectrum estimates with a significant resolution gap between the high-resolution RGB camera and low-resolution hyperspectral camera.

2. Related work

The spectrum can be densely measured using diffraction gratings, prisms, color filters or changing illumination. Du

*This work was done while the first author was visiting Microsoft Research Asia.

et al. used a prism and a mask to capture spectral images at video rate [8]. Color filters with a black/white camera are also widely used. The liquid crystal tunable filter [9] is popular since it has narrow passband that can be electronically adjusted. Schechner *et al*. [28] used a linear variable filter that has varying passband in one filter dimension. A custom-made linear CCD with changing color filters are used in [24]. A spectral video camera using two sets of CCD with different spectral filters are proposed in [32].

Illumination based methods modulate the light sources so that they emit lights of a specific wavelength. Zuzak *et al*. [34] used broadband light diffracted with digital micromirror device (DMD) to select which wavelength to be on or off. Park *et al*. [20] used five types of LEDs and an RGB camera to capture multi-spectral video.

It is not trivial to increase the spatial resolution of those cameras, and some of them are limited in the spectral resolution. Higher resolution in sensor itself would not solve the problem at once, since it can further cut down the number of photons reaching to the sensors, though the high-resolution in the spectral bandwidth already reduces them.

Motivated by the difficulty of building higher resolution hyperspectral sensors, researchers in remote sensing and color research have proposed a number of methods to fuse a high-resolution panchromatic image with a low-resolution hyperspectral image [6, 12, 2]. One popular technique first applies a linear transformation of color coordinates (*e.g*. HSI) to separate chromaticity and luminance into different channels. Since our human eyes are more sensitive to the luminance, this technique replace the luminance by the one from the high-resolution panchromatic image and then reverse the linear transformation to reconstruct the high-resolution multi-spectral images. These methods, however, tend to either distort the spectrum or blur the result since the values in chromaticity channels are not upsampled.

Another class of methods use image filtering techniques to interpolate the spectra in higher resolution [14]. Since these methods assume the interpolated values are spatially smooth, they often over-smooth the image details in the reconstructed hyper-/multi-spectral images. In contrast, our method respects the relationship between the RGB and hyperspectral channels, ensuring that the recovered high-resolution hyperspectral image produces the same high-resolution RGB image as the input. It also ensures that the more detailed spectral properties of the recovered image agree with those of the hyperspectral input.

Our approach accomplishes this by drawing on tools developed in the sparse representation community [4]. Conceptually, these tools are a good fit for our problem, since their goal sense signals and images using small (seemingly insufficient) sets of linear measurements. This seemingly impossible task is made possible by introducing strong prior information, in the form of the assumption that the signal is sparse in a given basis [25]. Naturally, to apply these techniques, we need to know that the signal is indeed sparse or *compressible* in a known basis, such as a wavelet basis. As we will see, in our problem, due to the special structure of the measurements, we may assume a much stronger form of sparsity: *each hyperspectral pixel images only a few materials*. However, this strong prior information comes with a price: unlike most applications of compressed sensing, here, the basis of sparsity, which corresponds to the reflectances of the materials in the scene is not known ahead of time, and must be estimated *from the observation itself* (similar to [15]).

Recently, a number of groups have applied similar techniques from sparse signal representation to hyperspectral data problems such as supervised classification [5] and hyperspectral unmixing [35, 17]. Our approach also makes use of hyperspectral unmixing as a key step, and uses an algorithm motivated by the sparsity of the signal in the spatial dimension. Our approach to unmixing has technical differences from the aforementioned works. For example, whereas [17] assumes sparsity in both the spectral and spatial domains, we only work with an assumption of spatial sparsity that follows naturally from the physical setting of the problem. However, the most important difference between the aforementioned works and our approach is that we utilized an inhomogeneous set of measurements: hyperspectral and RGB, and these different measurements have very different spatial sampling rates.

In the context of this paper, the most relevant related work is [23], which investigates the utility of wavelet sparsity priors for hyperspectral acquisition. Similar to our setting, this paper also augments the compressive hyperspectral measurements with an RGB image. However, as we will see, this work differs from ours in several respects: first, it requires a special hardware setup to obtain compressive hyperspectral measurements. Second, obtaining compressive measurements that multiplex over space may obscure a very natural spatial sparsity that arises due to the presence of a limited number of materials in the scene. We anticipate that for scenes in which the materials vary slowly relative to the sampling rate of the hyperspectral image, our method will be able to obtain a more accurate reconstruction by directly exploiting this sparsity.

3. Problem formulation

The goal of this work is to acquire a high-resolution spectral signal $\mathbf{Z} \in \mathbb{R}^{W \times H \times S}$, where W, H and S are the image width, the image height, and the number of sampled wavelengths respectively. Our system takes as input two types of measurements: An RGB image with a high spatial resolution $\mathbf{Y}_{rgb} \in \mathbb{R}^{W \times H \times 3}$, and a lower-resolution hyperspectral image $\mathbf{Y}_{hs} \in \mathbb{R}^{w \times h \times S}$, where $w \ll W$ and $h \ll H$. Both of these inputs are linear functions of the target signal:

$$\mathbf{Y}_{hs} = \mathcal{P}_{hs}[\mathbf{Z}], \quad \mathbf{Y}_{rgb} = \mathcal{P}_{rgb}[\mathbf{Z}], \tag{1}$$

where $\mathcal{P}_{hs} : \mathbb{R}^{W\times H\times S} \rightarrow \mathbb{R}^{w\times h\times S}$ and $\mathcal{P}_{rgb} : \mathbb{R}^{W\times H\times S} \rightarrow \mathbb{R}^{W\times H\times 3}$ are linear mappings.[1]

Since in general the total number of observations obtained from the high-resolution RGB and low-resolution hyperspectral images is smaller than unknowns $whS + 3WH \ll WHS$, the problem is highly under-constrained - reconstructing $\mathbf{Z}$ is impossible without further assumptions.

The first is that the number M of distinct materials in the scene is relatively small. Let $\mathbf{r}_j \in \mathbb{R}^S$ denote the reflectance function of the j-th material. If the illumination is uniform, then the radiance $\mathbf{a}_j$ of the j-th material will be the same at all points containing that material. Hence, at any point (i, j) in space, the spectrum $\mathbf{Z}(i, j, *) \in \mathbb{R}^S$ satisfies

$$\mathbf{Z}(i,j,*) \approx \sum_{m=1}^{M} \mathbf{a}_m \mathbf{h}_m(i,j) \doteq \mathbf{A}\mathbf{h}(i,j), \tag{2}$$

where $\mathbf{a}_m \in \mathbb{R}^S$ is the m-th radiance corresponds to the reflectance of the m-th material, and $\mathbf{h}(i, j) = [h_1(i, j), \ldots, h_M(i, j)]$ are the coefficients.

Now, each observation $\mathbf{Y}_{hs}(i, j)$ can be written as

$$\begin{aligned} \mathbf{Y}_{hs}(i,j,*) &\propto \sum_{k,l\in W_{ij}} \mathbf{Z}(l,k,*) \\ &= \mathbf{A} \sum_{k,l\in W_{ij}} \mathbf{h}(k,l) \doteq \mathbf{A}\mathbf{q}(i,j), \end{aligned} \tag{3}$$

where W_{ij} is an appropriate window of the high-resolution image, and $\mathbf{q}(i, j) \in \mathbb{R}^M$ are the summed coefficients. Second, this method assumes that at each pixel (i, j) of the low-resolution $\mathbf{Y}_{hs}$, only a few of the materials are present. Hence, the vector $\mathbf{q}(i, j)$ is a *sparse* vector - only a few of its entries are nonzero. This assumption is valid whenever the materials in the scene are not changing too quickly relative to the spatial sampling frequency of $\mathbf{Y}_{hs}$.

4. Solution via ℓ^1-minimization

If we concatenate the equations (3) for all pixels (i, j), this yields a matrix equation

$$\tilde{\mathbf{Y}}_{hs} = \mathbf{A}\mathbf{Q}, \tag{4}$$

where $\tilde{\mathbf{Y}}_{hs} \doteq [\mathbf{Y}_{hs}(1, 1, *), \ldots, \mathbf{Y}_{hs}(w, h, *)] \in \mathbb{R}^{S\times wh}$, and $\mathbf{Q} \doteq [\mathbf{q}(1, 1), \ldots, \mathbf{q}(w, h)] \in \mathbb{R}^{M\times wh}$. Eq. (4) expresses the hyperspectral observations as a product of two physically meaningful terms: The basis $\mathbf{A}$ of reflectance functions, and the sparse coefficients $\mathbf{Q}$ which represent the fractions of each material present at each location. These two matrices are initially unknown. However, for scenes that obey our above assumptions, we *do* know that the matrix $\mathbf{Q}$ should be sparse; ideally, as sparse as possible.

[1] Concretely, an RGB camera multiplies sensitivity functions $f_c(\lambda)$ to the incoming radiane $z(\lambda)$ and integrates over the spectral range Ω: $\int_\Omega f_c(\lambda)z(\lambda)d\lambda$, where $c = \{r, g, b\}$. In the matrix form, $\mathbf{Y}_{rgb} = \mathbf{F}\mathbf{Z}$, where $\mathbf{F}$ is a $3 \times S$ sensitivity matrix and $\mathbf{Z}$ is a $S \times WH$ matrix.

This is fortuitous: Although there are many possible decompositions of $\tilde{\mathbf{Y}}_{hs}$ as a product $\mathbf{A}\mathbf{Q}$, under fairly mild circumstances the *sparsest* such decomposition is unique.[2] Physically, such a decomposition attempts to explain each observation in $\mathbf{Y}_{hs}$ using the fewest number of materials possible. We seek such a decomposition by locally minimizing the ℓ^1 norm (sum of magnitudes) of the coefficients subject to a data constraint and a constraint on the ℓ^2 norm of the columns of $\mathbf{A}$:

$$\min_{\mathbf{A},\mathbf{Q}} \|\mathbf{Q}\|_1 \;\text{ s.t. }\; \tilde{\mathbf{Y}}_{hs} = \mathbf{A}\mathbf{Q},\; \|\mathbf{A}\mathbf{e}_i\|_2 = 1 \;\forall i. \tag{5}$$

This approach has been suggested and analyzed, *e.g.*, in [11].[3] We solve Eq. (5) via a non-smooth Gauss-Newton algorithm that repeatedly linearizes the nonlinear constraints and solves a sequence of ℓ^1-norm minimization problems. As a more detailed description is beyond the scope of this paper, we refer the interested reader to [10].

Once the basis $\mathbf{A}$ is recovered, we reconstruct $\mathbf{Z}$ using sparsity as a guide. Assuming $\mathbf{A}$ has been correctly estimated, each pixel of the target high-resolution image should admit a sparse representation in terms of the columns of $\mathbf{A}$: $Z(i, j, *) = \mathbf{A}\mathbf{h}(i, j)$. We seek the sparsest coefficients $\mathbf{h}(i, j)$ that satisfy the measurement equation $\mathbf{Y}_{rgb}(i, j, *) = \mathcal{P}_{rgb}\mathbf{A}\mathbf{h}(i, j)$:

$$\begin{aligned} \mathbf{h}(i,j) \;&=\; \operatorname{argmin}_{\mathbf{h}} \|\mathbf{h}\|_1 \\ \text{s.t.} \quad &\quad \|\mathbf{Y}_{rgb}(i,j,*) - \mathcal{P}_{rgb}\mathbf{A}\mathbf{h}\|_2 \le \varepsilon. \end{aligned} \tag{6}$$

Above, ε is a parameter that reflects the noise level in the RGB image. We can then use the estimated coefficients $\hat{\mathbf{h}}$ to form our estimate of $\mathbf{Z}$:

$$\hat{\mathbf{Z}}(i,j,*) = \mathbf{A}\hat{\mathbf{h}}(i,j). \tag{7}$$

5. Experiments

To assess the performance of the proposed method, we conduct both simulation and real-world experiments. We compare the proposed method with the component substitution method [2] as well as three methods based on Principal Component Analysis (PCA) [22] that we have developed. In Section 5.2, we use these methods as baselines for a quantitative comparison with real hyperspectral data and a simulated RGB image. Finally, in Sec. 5.3, we show the effectiveness of our method in practice, using images acquired with our own hardware setup.

[2] For example, if $n \le S$ and $\mathbf{Q}$ has Bernoulli nonzeros with magnitudes following from any absolutely continuous probability distribution, it is not difficult to show that with high probability the decomposition is unique provided $wh = \Omega(n \log n)$.

[3] The reader familiar with the sparse approximation literature may recognize the problem of factoring $\tilde{\mathbf{Y}}_{hs}$ into $\mathbf{A}$ and $\mathbf{Q}$ as an instance of the "dictionary learning" problem discussed in the survey paper [26] and references therein. Study of this problem in imaging dates back at least to the seminal work of Olshausen and Field [19], since which many local optimization algorithms have been proposed. We expect that many of these algorithms, such as the K-SVD [1], can also be profitably applied to our problem.

Balloons | Beads | Sponges | Oil painting | Flowers | CD | Fake and real peppers | Photo and face

Figure 1. Simulated RGB images using the database [33].

Method	Image set							
	Balloons	Beads	Sponges	Oil painting	Flowers	CD	Fake and real peppers	Photo and face
CSM [2]	13.9	28.5	19.9	12.2	14.4	13.3	13.7	13.1
Global	6.9/4.7	10.5/8.8	15.4/12.3	5.4/**3.8**	9.8/8.9	10.3/10.0	7.1/5.9	4.7/3.8
Local window	7.0/4.9	10.6/8.9	14.0/10.6	5.7/4.1	7.5/6.3	9.6/9.2	8.8/8.0	10.9/10.5
RGB clustering	6.6/4.3	9.7/**7.9**	13.6/10.0	5.5/4.0	7.8/6.5	9.1/8.6	8.5/7.6	4.7/3.8
Factorization	**3.0/3.0**	**9.2**/9.2	**3.7/3.7**	**4.7** / 4.7	**5.4/5.4**	**8.2/8.2**	**4.7/4.7**	**3.3/3.3**

Table 1. Numerical comparison of RMSE (Root Mean Squared Error) between methods. Two numbers are before/after the back-projection refinement. The numbers are shown in the range of 8-bit images. Those in bold are the best score among the methods.

5.1. Baselines: PCA Variants

The approaches in this section are motivated by the empirical observation that real scene radiances vary smoothly with wavelength, and hence can be represented using relatively small numbers of basis functions [16, 30, 7, 21]. One natural approach to obtaining basis is to apply Principal Component Analysis (PCA) to the hyperspectral observations.

Let $\mathbf{A}_{\text{PCA}} \in \mathbb{R}^{S\times 3}$ be a matrix whose columns are the first three principal components of $\bar{\mathbf{Y}}$. If the columns of $\mathbf{A}_{\text{PCA}}$ capture most of the variance in the hyperspectral image, and they are not orthogonal to $\mathcal{P}_{rgb}$, we may be able to use them to estimate $\mathbf{Z}$, via

$$\mathbf{q}(i,j) = (\mathcal{P}_{rgb}\mathbf{A}_{\text{PCA}})^{\dagger}\mathbf{Y}_{rgb}(i,j,*), \tag{8}$$

$$\hat{\mathbf{Z}}_{\text{PCA}}(i,j,*) = \mathbf{A}_{\text{PCA}}\mathbf{q}(i,j), \tag{9}$$

where $(\cdot)^{\dagger}$ denotes the matrix pseudo-inverse. We call this approach the "Global PCA" method.

Since the variation of spectrum across the whole scene can be larger than three dimensions, the straightforward global method might not accurately reconstruct the hyperspectral image. In this case, it may be better to replace the global linear model of PCA by a local linear model, given by fitting data in a neighborhood of each given pixel. Because these local windows overlap, the estimated local bases share some spatial coherence. We call this method a "Local window" method in this experiment.

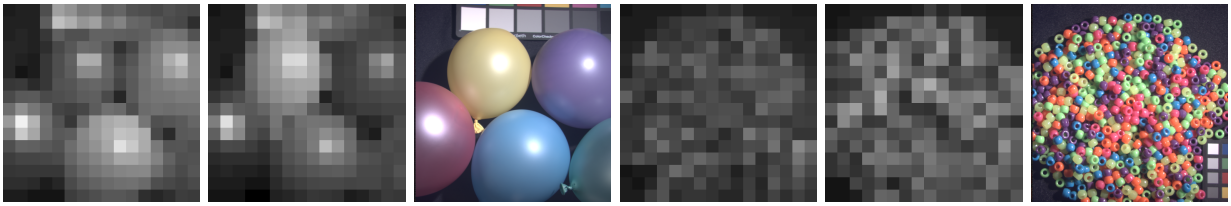
The two methods based on local window and the global PCA method have individual drawbacks. The global PCA method does not have flexibility to represent variations in spectra, and the local window method does not consider any color similarity in each window. To include both local and global characteristics, another idea is to cluster the high-resolution pixels based on their RGB values, and then estimate a separate basis for each color cluster individually. This "RGB clustering" method assumes that for the same scene with the same RGB color value, it should produce the same hyperspectral values. Although this assumption is limited in some situations, we find that it leads to better estimates than simple PCA methods.

Refinement by back-projection If the reconstructed image $\hat{\mathbf{Z}}$ does not satisfy the observation equation $\mathcal{P}_{rgb}[\hat{\mathbf{Z}}] = \mathbf{Y}_{rgb}$, it can be further refined using via backprojection, an interative method for finding a minimum-norm perturbation of $\hat{\mathbf{Z}}$ such that the observation equation is satisfied. This approach is often used in image super-resolution methods [29]. In image super-resolution, the reconstructed high-resolution image, after down-sampling, needs to be the same as the original low resolution input. In our case, we can also incorporate the second constraint $\mathcal{P}_{hs}[\mathbf{Z}] = \mathbf{Y}_{hs}$ to further improve the result. In our experiments, we find that for the simple PCA methods introduced above, backprojection bring $\mathbf{Z}$ significantly closer to the ground truth.

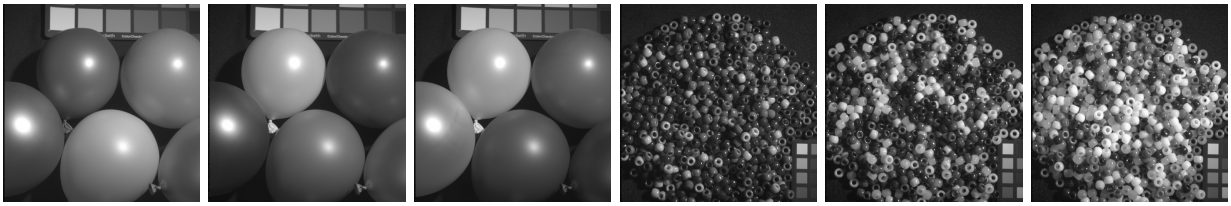
5.2. Performance and Accuracy

We have conducted experiments using a spectral image database described in [33], which provides spectral images of 32 scenes, sampled every 10 nm from 400 nm to 700 nm, by using a liquid crystal tunable filter, VariSpec [31] with a black/white cooled CCD camera, Apogee Alta U260 [3]. Images were taken with fixed focus, in 16-bit PNG format, and the resolution is 512 by 512 pixels.

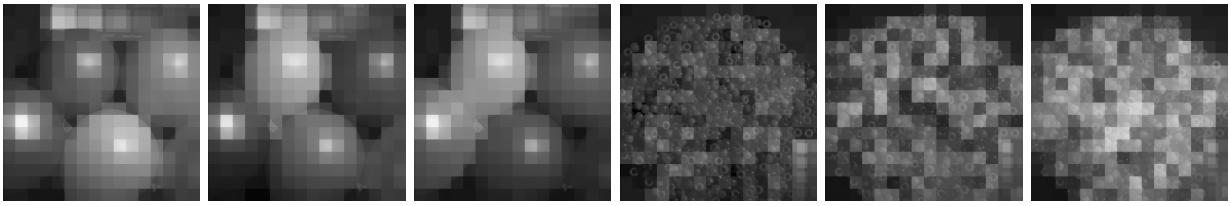
To simulate a low-resolution camera, we down-sampled them to 16 by 16 pixels. The RGB camera response were simulated by integrating them over the wavelength, using the filter characteristics of DSLR camera, Nikon D1. We



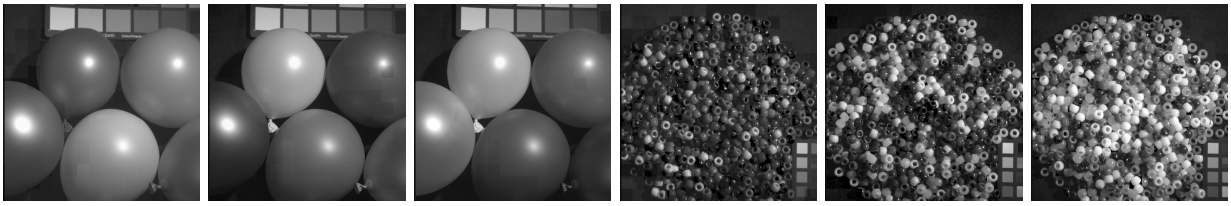
Input images: Balloons and Beads examples.



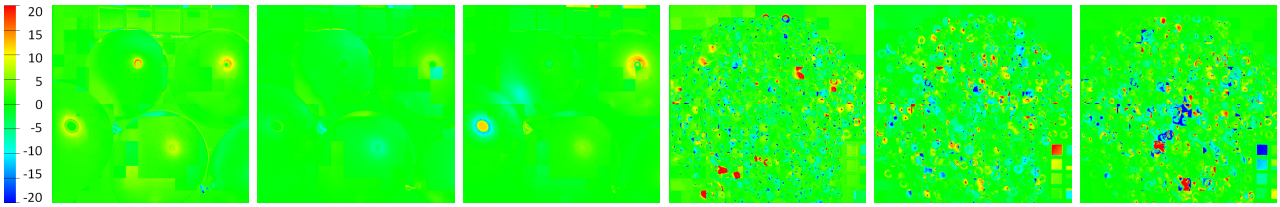
Ground truths.



Reconstruction using component substitution method [2].



Reconstruction using matrix factorization and refinement.



Error images of the matrix factorization and refinement.

Figure 2. Spectral images. The top row: Input images, sampled at 460, 550 nm, and the RGB image, from left to right. The first three images are from Balloons example. The remaining images are the inputs of Beads. The second row: Ground truths, sampled at 460, 550, and 620 nm. The order is the same in the following rows. The third row: Reconstructed results of component substitution method [2]. The fourth row: Results of the proposed method. The bottom row: Error images of the proposed method.

used 8 examples in the database, and their simulated RGB images are shown in Fig. 1. Down-sampled images are shown in the top row in Fig. 2.

To evaluate the performance of the proposed method, we implemented the following five methods: (1) Component substitution method [2], (2) Global PCA, (3) Local window, (4) RGB clustering methods (described in Sec. 5.1), and the proposed method - (5) Matrix factorization method (described in Sec. 4.) Then, we apply a refinement method to the initial estimates from these methods by the back-projection described in Sec. 5.1 For the RGB clustering, we used k-means clustering with $k = 10$, and used chromaticity and spatial coordinate for the feature space. For the matrix factorization method, we used 0.5 for ε in Eq. (6), and used the top 6 bases in A that contribute to each low-resolution spectrum, when doing the reconstruction.

The RMSE (Root Mean Squared Error) between the reconstructed spectral images and the ground truths (original data) are summarized in Table 1. Each row shows the RMSE of the initial estimate of each method, and the RMSE

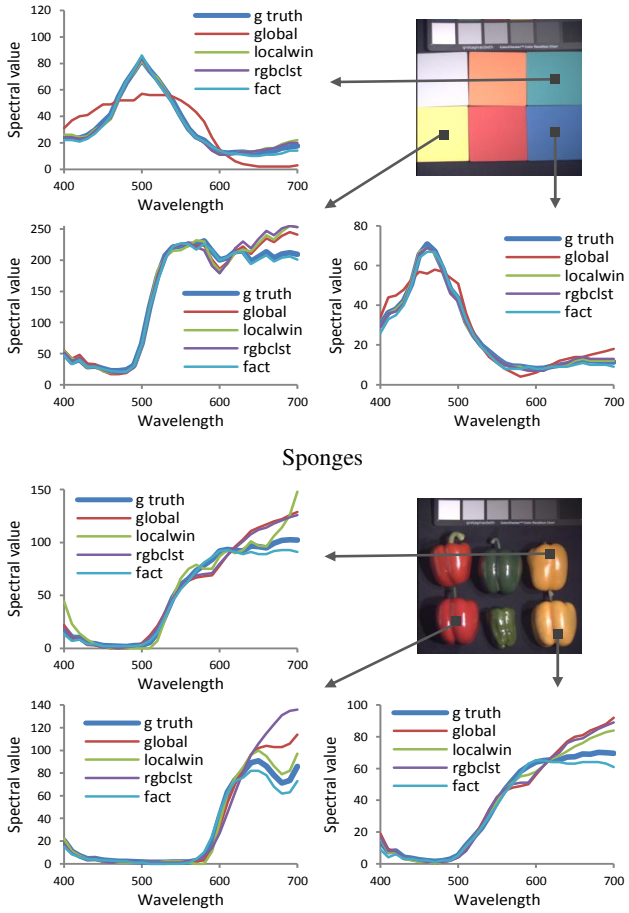


Figure 3. Comparison of spectra.

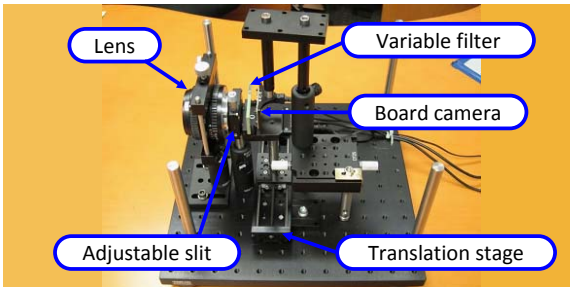


Figure 4. Hyperspectral camera.

after the back-projection refinement. The RMSE values are shown in the range of 8-bit images. The average of the RMSEs are 16.1, 7.3, 7.8, 6.6 and 5.3 (6.3, 2.9, 3.1, 2.6, and 2.1 %) of 255 for CSM, Global PCA, Local window, RGB clustering and Matrix factorization methods. The numbers show the effectiveness of the proposed method. The table also shows that back-projection further improves the reconstruction error of the initial estimate.

The examples of the reconstructed images are shown in Fig. 2. The first and second rows show the inputs and the ground truths of Balloons and Beads examples, of 460, 550, and 620 nm. The third row shows the reconstructed results of the component substitution method [2], which suffers from the large resolution difference between the hyperspectral and RGB images. The fourth and the bottom rows show the results and the errors of the matrix factorization method. The positive errors are shown in red, and negative ones are in blue. The proposed method succeeds to recover the details while maintaining the spectral information. The errors are large around specular pixels that changes rapidly with respect to the image resolution.

Fig. 3 shows the reconstructed spectra of three pixels in Sponges and Fake and real peppers. The ground truth is shown as a bold line, and the reconstructed spectra of (1) Global PCA, (2) Local window, (3) RGB clustering and (4) Matrix factorization are also shown. Global PCA tends to distort the spectra due to the gross assumption that every spectra can be represented by three dimensional basis functions. All the methods perform reasonably well, but above all, the matrix factorization method provides the best results.

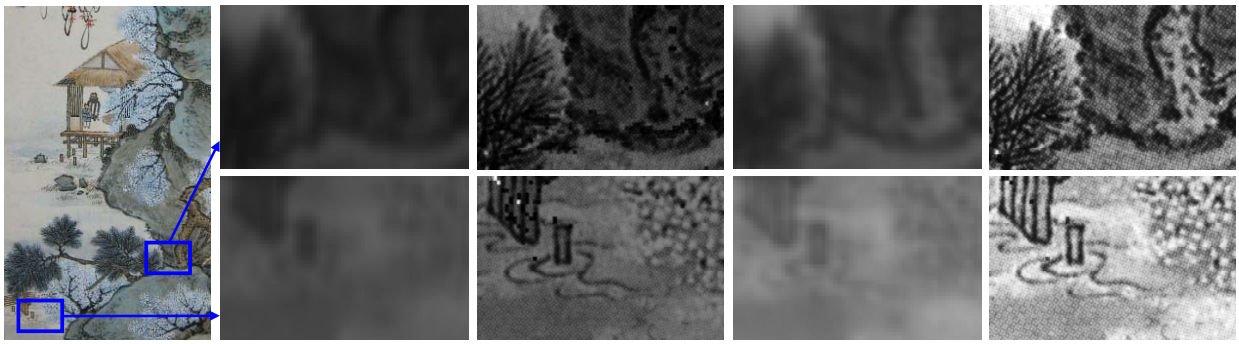
5.3. Real-world examples

Experimental setup In order to obtain real image data to validate our hybrid frame-work we need two types of images. The first is high-spatial resolution RGB camera with known spectral and intensity response functions, and the second is low spatial resolution, hyperspectral camera.

To obtain high-resolution RGB images, we used the 4008×2672 Lumenera LW11509 camera with a long focal length lens. This camera uses the Kodak KAI-11002 sensor that has large pixel, linear response function and known spectral response for the RGB channels. The camera was used with an integral Schott BG40 as an IR cutoff filter.

To obtain hyperspectral images, we built the device shown in Fig. 4. This device uses variable linear band pass filter (Schott Veril 60), in conjunction with an adjustable slit (Thorslab VA100/M) for bandwidth control and a CMOS B/W sensor with a known response function (Point Grey Firefly MV set to 640×480 resolution). The position of the filter is controlled by a computerized translation stage (Zaber zaber KT-LSM100A). This setup is more economical than a tunable filter that are significantly more expensive, and provides better control than uncontrolled modulation imaging [27] because the filter is stationary during image capture at any desired exposure time. We calibrated our device using several narrow band filters. Because the repeatability of the translation stage is very accurate, we only need to calibrate once. The camera can sample spectra at every 5 nm from 415 nm to 730 nm. We also used several narrow band filters to capture short wavelengths under 450 nm to ensure reliable measurement.

To capture images from the same view point, a beam splitter is usually used. However, since we did not need to capture them simultaneously, we simply replaced the cam-



Input RGB | Input (550nm) | Estimated (550 nm) | Input (620 nm) | Estimated (620 nm)

Figure 5. The figures show two zoomed views of the high-resolution image: The input and estimated spectral images of 550 and 620 nm.

era while maintaining the same view point. The slight difference of viewpoints were compensated by using the affine transformation.

Results The input and reconstructed images of a printed painting is shown in Fig. 5. The figure shows two zoomed views of the painting. The high resolution RGB camera can visualize the printing pattern, which is neither visible to the human eyes nor to the low-resolution camera. The proposed method recovers spectral images of the painting in details, and provides visually pleasing results as shown in the figure. With the proposed method, highly detailed spectral information is acquired in a very efficient manner, both in terms of the acquisition time and the cost of the equipment.

6. Discussion

As shown in the experiments, the matrix factorization method and the RGB clustering method perform better than the others. This may not be surprising, since the two methods are similar in a sense that they utilize a kind of segmentation in spectra/RGB domains. They may be further improved by jointly using those sources of information. The distribution of the reconstruction errors of the proposed method is unique compared to those of the PCA-based methods. The proposed method sometimes fails to reconstruct specific colors, while errors are equally distributed to many colors when PCA-based methods were used. PCA-based methods seem to suffer more than the proposed method, because they discard the spectral information that lies in more than three dimensions.

The proposed method has more advantages in contrast with interpolation/super-resolution or pan-sharpening methods, since these methods can usually only magnify an image up to a factor of 8, but our proposed method can magnify the hyperspectral image 32 times larger (32×32 resolution). Images with such a large magnification factor are almost impossible to be reconstructed correctly by the current state-of-the-art methods.

There are limitations with the proposed method. The method sometimes produces errors in some bands, when more than *a few* number of basis functions are required to reconstruct the full spectral image. This is a general limitation for all methods that assume a small fixed number of basis functions. However, in scenes that are composed of a limited number of materials, our results clearly surpass the results of all other straightforward methods. Thus, our method has an advantage over these methods.

The refinement by the back-projection was not significant for the results of the matrix factorization method. This is because the downsampled values from the recovered spectra are already very close to those in the original RGB/spectral images when using the method.

The issue with the matrix factorization method is the computational time. For 4008×2672 images, it takes several hours to do the factorization, and several hours to do the reconstruction, using a high-end machine with eight cores. The current implementation uses Matlab 7.0. We plan to speed up the process by using GPU or multi-thread programming. There are parameters that have to be chosen with the method. The epsilon depends on the noise level, and the number of bases that contribute to a pixel is chosen empirically in the paper, but is fixed for all examples. We consider that the latter can be roughly determined as long as the convergence of the factorization is ensured.

7. Conclusion

We have proposed an efficient method to acquire high-resolution hyperspectral images, by jointly using a high-resolution RGB camera with a low-resolution hyperspectral camera. Assuming that only a few materials contribute to each pixel in the hyperspectral image, our method finds the optimal set of basis functions by factorizing the hyperspectral image into the basis and its sparse coefficients via ℓ^1-norm minimization. The spectrum at high-resolution is then reconstructed using the estimated basis and the RGB values. Comparison between the method and several simple approaches shows the effectiveness of the method. The

method enables to capture high-resolution hyperspectral image at reasonable cost and acquisition time.

Acknowledgement

This work was in part supported by Microsoft Research CORE 6 Project 2010.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. The K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. on Signal Processing*, 51(11):4311–4322. 2331

[2] B. Aiazzi, S. Baronti, and M. Selva. Improving component substitution pansharpening through multivariate regression of ms+pan data. *IEEE Trans. Geoscience and Remote Sensing*, 45(10):3230–3239, 2007. 2330, 2331, 2332, 2333, 2334

[3] Apogee. http://www.ccd.com/alta_u260.html. 2332

[4] A. Bruckstein, D. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009. 2330

[5] A. Castrodad, Z. Xing, J. Greer, E. Bosch, L. Carin, and G. Sapiro. Discriminative sparse representations in hyperspectral imagery. *IMA Preprint #2302, University of Minnesota*, 2010. 2330

[6] P. S. Chavez, S. C. Sides, and J. A. Anderson. Comparison of three different methods to merge multiresolution and multispectral data: Landsat tm and spot panchromatic. *Photo. Eng. Rem. Sens.*, 57(3):295–303, 1991. 2330

[7] J. Cohen. Dependency of the spectral re ectance curves of munsell color chips. *Psychon. Sci.*, 1:369–370, 1964. 2332

[8] H. Du, X. Tong, X. Cao, and S. Lin. A prism-based system for multi-spectral video acquisition. In *ICCV*, 2009. 2330

[9] S. C. Gebhart. *Liquid-crystal tunable filter spectral imaging for discrimination between normal and neoplastic tissues in the brain*. PhD thesis, Vanderbilt University, 2006. 2330

[10] Q. Geng, H. Wang, and J. Wright. On the local correctness of ℓ^1-minimization for dictionary learning. preprint, available at http://arxiv.org/abs/1101.5672, 2011. 2331

[11] R. Gribonval and K. Schnass. Dictionary identification: Sparse matrix factorization via ℓ^1-minimization. *IEEE Transactions on Information Theory*, 56(7), 2010. 2331

[12] F. H. Imai and R. S. Berns. High-resolution multi-spectral image archives: a hybrid approach. In *Color Imaging Conference*, pages 224–227, 1998. 2330

[13] N. Keshava and J. Mustard. Spectral unmixing. *Signal Processing Magazine*, 19(1):44–57, 2002. 2329

[14] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. *ACM Trans. Graphics*, 26(3):96–1–96–5, 2007. 2330

[15] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *ICCV*, 2009. 2330

[16] L. T. Maloney and B. A. Wandell. Color constancy: a method for recovering surface spectral reflectance. *JOSA*, 3(1):29–33, 1986. 2332

[17] Y. Moudden, J. bobin, J. Starck, and J. Fadili. Blind source separation with spatio-spectral sparsity constraints – application to hyperspectral data analysis. In *Proc. Int'l Conf. on Ind. Comp. Anal. Sig. Sep.*, 2009. 2330

[18] H. V. Nguyen, A. Banerjee, and R. Chellappa. Tracking via object reflectance using a hyperspectral video camera. In *Proc. CVPR Workshop*, 2010. 2329

[19] B. Olshausen and D. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37:3311–3325, 1997. 2331

[20] J.-I. Park, M.-H. Lee, M. D. Grossberg, and S. K. Nayar. Multispectral imaging using multiplexed illumination. In *ICCV*, 2007. 2330

[21] J. P. S. Parkkinen, J. Hallikainen, and T. Jaaskelainen. Characteristic spectra of munsell colors. *JOSA*, 6(2):318–322, 1989. 2332

[22] K. Pearson. On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, 2:559–572, 1901. 2331

[23] Y. Pfeffer and M. Zibulevsky. A micro-mirror array based system for compressive sensing of hyperspectral data. *Technion Technical Report 2010-10*, 2010. 2330

[24] A. Ribes, F. Shmitt, R. Pillay, and C. Lahanier. Calibration and spectral reconstruction for crisatel: an art painting multispectral acquisition system. *J. Im. Sci. Tech.*, 49:563–573, 2005. 2330

[25] J. Romberg. Imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–31, 2008. 2330

[26] R. Rubenstein, A. Bruckstein, and M. Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):972–982, 2010. 2331

[27] Y. Schechner and S. Nayar. Uncontrolled Modulation Imaging. In *CVPR*. 2334

[28] Y. Y. Schechner and S. K. Nayar. Generalized mosaicing. In *ICCV*, 2001. 2330

[29] J. Sun, N. N. Zheng, H. Tao, and H. Y. Shum. Image hallucination with primal sketch priors. In *CVPR*, 2003. 2332

[30] D. Y. Tzeng and R. S. Berns. A review of principal component analysis and its applications to color technology. *Color Res. Appl.*, 30(2):84–98, 2005. 2332

[31] Varispec. http://www.cri-inc.com/products/varispec.asp. 2332

[32] M. Yamaguchi, H. Haneishi, and N. Ohyama. Beyond red-green-blue (rgb): Spectrum-based color imaging technology. *J. Im. Sci. Tech.*, 52(1):010201–010201–15, 2008. 2330

[33] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar. Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum. *IEEE Trans. IP*, 19(9):2241–2253, 2010. 2332

[34] K. J. Zuzak, R. P. Francis, E. F. Wehner, J. Smith, M. Litorja, D. W. Allen, C. Tracy, J. Cadeddu, and E. Livingston. Dlp hyperspectral imaging for surgical and clinical utility. In *Proc. SPIE*, 2009. 2330

[35] A. Zymnis, S. Kim, J. Skaf, M. Parente, and S. Boyd. Hyperspectral image unmixing via alternating projected subgradients. In *Proc. of Asilomar Conference on Signals, Systems and Computers*, pages 1164–1168, 2007. 2330