

# Hyperspectral Super-Resolution by Coupled Spectral Unmixing

Charis Lanaras, Emmanuel Baltsavias, Konrad Schindler  
Institute of Geodesy and Photogrammetry, ETH Zürich, Switzerland

firstname.lastname@geod.baug.ethz.ch

## Abstract

Hyperspectral cameras capture images with many narrow spectral channels, which densely sample the electromagnetic spectrum. The detailed spectral resolution is useful for many image analysis problems, but it comes at the cost of much lower spatial resolution. Hyperspectral super-resolution addresses this problem, by fusing a low-resolution hyperspectral image and a conventional high-resolution image into a product of both high spatial and high spectral resolution. In this paper, we propose a method which performs hyperspectral super-resolution by jointly unmixing the two input images into the pure reflectance spectra of the observed materials and the associated mixing coefficients. The formulation leads to a coupled matrix factorisation problem, with a number of useful constraints imposed by elementary physical properties of spectral mixing. In experiments with two benchmark datasets we show that the proposed approach delivers improved hyperspectral super-resolution.

## 1. Introduction

Hyperspectral cameras sample the electromagnetic spectrum in many contiguous, very narrow spectral bands. Each pixel thus observes a detailed distribution of the surface reflectance across the visible (and near infrared) spectrum, rather than just three or four broad “colour” channels (see Fig. 1). Each physical material has a characteristic reflectance spectrum, therefore hyperspectral sensors (a.k.a. imaging spectrometers) can more accurately distinguish visually similar surface materials. This capability has proved useful for visual tasks such as tracking [23] and segmentation [22], in applications including face recognition [19], document analysis [18], food inspection [26] and earth observation [6]. However, the high spectral resolution comes at a cost. The narrow slicing of the spectrum means that in each band only a small fraction of the overall radiant energy reaches the sensor. To achieve an acceptable signal-to-noise ratio one must therefore increase the pixel size on the chip, and consequently the pixel footprint on the surface (or one

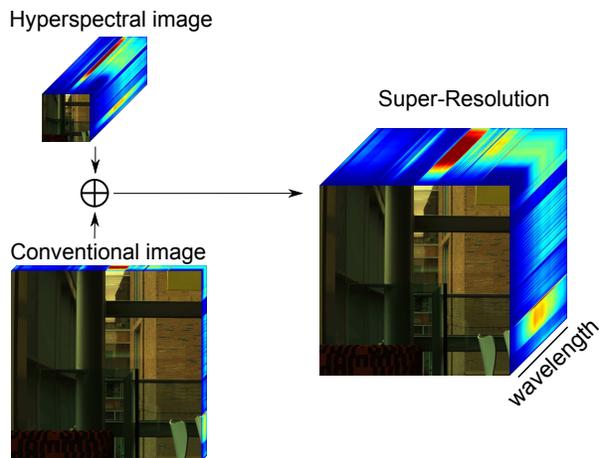


Figure 1. Hyperspectral images have high spectral resolution but low spatial resolution, whereas the opposite is true for conventional images. Hyperspectral super-resolution aims to fuse the two types of imagery.

has to increase integration time, which is impractical in the presence of camera or object motion because of the ensuing motion blur). Therefore hyperspectral images have significantly lower spatial resolution compared to conventional RGB cameras (or similar broadband sensors such as near-infrared cameras).<sup>1</sup> Conventional cameras on the other hand lose much of the spectral information, but by integrating the scene radiance over broad spectral bands they achieve high spatial resolution.

At hardware level the trade-off between detailed spatial and spectral information is due to fundamental physical limits and will thus be difficult to overcome. A natural solution is to instead record a hyperspectral and a conventional image, and to fuse them into a product with high spatial and spectral resolution, as illustrated in Fig. 1. This procedure is referred to as *hyperspectral image fusion* or *hyperspectral super-resolution*.

<sup>1</sup>In remote sensing, cameras with broad spectral bands are collectively referred to as “multispectral”. Unfortunately, in other fields “multispectral” is sometimes used for images with *higher* spectral resolution than RGB; we thus avoid the term.

The starting point for the present paper is that hyperspectral super-resolution is intimately related to another problem, namely *hyperspectral unmixing*: due to the much larger pixel footprint, hyperspectral images tend to have a large number of so-called “mixed pixels”, where the spectral responses of several different surface materials are overlaid. Unmixing tries to separate those responses, i.e. recover the underlying pure material spectra (called *endmembers*) and their relative proportions (called *abundances*). Unmixing is in itself an important processing step often needed for further analysis (e.g. semantic segmentation into different materials for surface inspection, ecological studies, geological prospection etc.). Beyond that, however, it also carries valuable information for the super-resolution problem, because the endmembers and their abundances in any given region must be preserved during upsampling.

**Contribution.** In this paper we develop a method which jointly solves hyperspectral super-resolution and unmixing. Given a low-resolution hyperspectral image and a high-resolution conventional image, we estimate a high-resolution hyperspectral image *and* a physically plausible decomposition of each pixel into spectral endmembers. It turns out that the joint formulation significantly improves super-resolution. Unmixing imposes physical constraints: in particular, neither the reflected intensity at any wavelength nor the surface area covered by a material can ever be negative, and each observed spectrum should be completely described by a mixture of few materials.

At the technical level we rely on the *linear mixing model* (LMM). We cast the problem as a coupled, constrained factorisation of the two input images into endmembers (pure spectral signatures) and abundances (mixing coefficients). Modern optimisation methods [8] allow for an efficient local optimisation of the resulting objective function.

Experiments on two widely used standard benchmarks show that the proposed approach delivers improved fusion results as well as a visually convincing soft segmentation into spectrally distinct materials.

## 2. Related Work

The limitation that hyperspectral images can only be acquired at low spatial resolution has naturally led researchers in computer vision and remote sensing to try fusing them with conventional high-resolution images, to obtain a synthetic product of high spatial and spectral resolution. From the point of view of image processing, the problem is a special case of image fusion [24]. In particular, it can be seen as a generalisation of pan-sharpening. In remote sensing [20] and colour research [14] pan-sharpening refers to the task of enhancing a lower-resolution colour image by fusing it with a single-channel black-and-white (“panchromatic”)

image of higher resolution. Pan-sharpening by and large relies on the observation that colour variations have a lower frequency than brightness changes (the same principle is also used in image compression). Image transformations such as for example HSV, Brovey or PCA are employed to isolate the intensity channel and replace it with the higher-resolution panchromatic image. For an overview see [3].

Another natural way to increase the resolution of hyperspectral images is to acquire multiple such images from slightly different viewpoints and apply standard super-resolution methods to them [31]. Note however, that such an approach in practice needs a lot more data than the fusion with a high-resolution image to reach good results. Also, it may be difficult to acquire multiple jittered images, e.g. in the case of satellite imaging or if the objects in the scene move. It has also been suggested to learn a prior for the local structure of the upsampled images from offline training data [11]. This is an instance of (learning-based) blind deblurring, and will at most work for moderate upsampling.

Our work is based on the linear mixing model, consequently it is most closely related to methods that also rely on a linear basis and some form of matrix factorisation. Kawakami *et al.* [16] proceed sequentially: they first learn a spectral basis by unmixing the hyperspectral image via  $l_1$ -minimisation. With that (fixed) basis they then compute mixing coefficients for the high-resolution RGB pixels, again using  $l_1$ -minimisation to enforce sparsity of the mixing weights. Similarly, Huang *et al.* [13] learn the spectral basis with SVD, and solve for the high-resolution mixing coefficients with orthogonal matching pursuit (OMP). Also along these lines, Akhtar *et al.* [1] learn a non-negative spectral basis from the hyperspectral image, and then solve for the high-resolution coefficients under a sparsity constraint, using OMP. Simões *et al.* [21] also find a linear basis, and include a total variation regularizer to favour spatial smoothness of the mixing coefficients. Other than the previously described approaches, Yokoya *et al.* [29] also update the initial basis. They emphasise that the coefficients must be positive and use non-negative matrix factorisation to unmix both the hyperspectral and the RGB image in a coupled fashion. Wycoff *et al.* [27] also formulate a joint energy function over both the basis vectors and the coefficients. The energy promotes sparsity and non-negativity of the basis and the coefficients. Bayesian approaches [12, 25] additionally impose priors on the distribution of the image intensities and do MAP inference, which under simple priors boils down to reweighting the contributions of different pixels to the error function. [15] also add a MRF prior to model spatial dependence. Akhtar *et al.* [2] employ non-parametric Bayesian dictionary learning to obtain a spectral basis, and then obtain the super-resolution image with Bayesian sparse coding.

In our work we also rely on the factorisation of a linear

mixture model, but attempt to base the super-resolution on an optimal, physically plausible reconstruction of the endmembers and their abundances. While most state-of-the-art work uses some of the constraints that result from the physics of spectral mixing, we are not aware of one that exploits all of them. To account for the influence of the constraints on the spectral basis, we update the endmembers together with the abundances, whereas [16, 13, 1, 21, 2] estimate the spectral basis in advance and then keep it fixed. Furthermore, we force the factorisation to correspond to a feasible mixture of material spectra: we require that the intensities are  $\geq 0$  for all wavelengths, that the mixing coefficients are  $\geq 0$  and sum to 1 (*i.e.*, proportions of different materials in a pixel add up), and that the mixture in every pixel is sparse. On the contrary, [21, 13, 2] construct an arbitrary subspace basis without a physical interpretation; [16, 2] does not enforce non-negativity; [27, 29, 1, 2] do not ensure the coefficients sum to 1, and [29, 21, 2] also ignore sparsity. We argue that a physically grounded unmixing constrains the problem more tightly, and thus will yield not only more plausible endmembers, but also a more correct super-resolution.

### 3. Problem Formulation

We are searching for an image  $\bar{\mathbf{Z}} \in \mathbb{R}^{W \times H \times B}$  that has both high spatial and high spectral resolution, with  $W$ ,  $H$  and  $B$  the image width, image height and number of spectral bands respectively. For that task we have two inputs: a hyperspectral image  $\bar{\mathbf{H}} \in \mathbb{R}^{w \times h \times B}$  with (much) lower spatial resolution, *i.e.* the same region in object space is covered by a smaller number of pixels:  $w \ll W$  and  $h \ll H$ ; and a conventional image  $\bar{\mathbf{M}} \in \mathbb{R}^{W \times H \times b}$  with high spatial resolution, but a reduced number of spectral bands,  $b \ll B$ .

To simplify the notation, we will write images as matrices, *i.e.* all pixels of an image are concatenated, such that every column of the matrix corresponds to the spectral responses at a given pixel, and every row corresponds to the complete image in a specific spectral band. Accordingly, the images are written  $\mathbf{Z} \in \mathbb{R}^{B \times N_m}$ ,  $\mathbf{H} \in \mathbb{R}^{B \times N_h}$  and  $\mathbf{M} \in \mathbb{R}^{b \times N_m}$ , where  $N_h = wh$  and  $N_m = WH$ .

According to the linear mixing model [17, 7] the intensities  $\mathbf{z} \in \mathbb{R}^B$  at a given pixel of  $\mathbf{Z}$  are described by an additive mixture

$$\mathbf{z} = \sum_{j=1}^p \mathbf{e}_j a_j \quad , \quad \mathbf{Z} = \mathbf{E} \mathbf{A} \quad , \quad (1)$$

with a matrix  $\mathbf{E} \equiv [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p]$  of endmembers and a matrix  $\mathbf{A} \equiv [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{N_m}]$  of per-pixel abundances. By this definition, at most  $p$  endmembers (materials) are visible in the image. The endmembers  $\mathbf{E}$  act as a non-orthogonal basis to represent  $\mathbf{Z}$  in a lower-dimensional space  $\mathbb{R}^p$ , and  $rank\{\mathbf{Z}\} \leq p$ .

The actually recorded low-resolution hyperspectral image  $\mathbf{H}$  is a spatially downsampled version of  $\mathbf{Z}$ ,

$$\mathbf{H} \approx \mathbf{Z} \mathbf{S} = \mathbf{E} \mathbf{A} \mathbf{S} = \mathbf{E} \tilde{\mathbf{A}} \quad , \quad (2)$$

where  $\mathbf{S} \in \mathbb{R}^{L_m \times L_h}$  is the downsampling operator that describes the spatial response of the sensor, and  $\tilde{\mathbf{A}} \equiv \mathbf{A} \mathbf{S}$  are the abundances at the lower resolution – under a linear downsampling simply the (weighted) average of the high-resolution abundances within one low-resolution pixel.

Similarly, the high-resolution conventional image  $\mathbf{M}$  is a spectrally downsampled version of  $\mathbf{Z}$ ,

$$\mathbf{M} \approx \mathbf{R} \mathbf{Z} = \mathbf{R} \mathbf{E} \mathbf{A} = \tilde{\mathbf{E}} \mathbf{A} \quad , \quad (3)$$

where  $\mathbf{R} \in \mathbb{R}^{b \times B}$  is the spectral response function of the sensor and  $\tilde{\mathbf{E}} \equiv \mathbf{R} \mathbf{E}$  are the spectrally degraded endmembers (for a standard camera the RGB values of different materials). The spatial response function  $\mathbf{S}$  of the hyperspectral camera and the spectral response function  $\mathbf{R}$  of the conventional camera form part of the camera specifications and are assumed to be known.

**Constraints.** The core idea of the present paper is to improve super-resolution by making full use of the fact that in (1) the endmembers are reflectance spectra of individual materials, and the abundances are proportions of those endmembers. As a consequence, the factorisation is subject to the following constraints:

$$\begin{aligned} a_{ij} &\geq 0 \quad \forall i, j && \text{(non-negative abundance)} \\ \mathbf{1}^\top \mathbf{A} &= \mathbf{1}^\top && \text{(abundances sum to 1)} \\ 0 &\leq e_{ij} \leq 1 \quad \forall i, j && \text{(non-negative, bounded reflectance)} \end{aligned} \quad (4)$$

with  $e_{ij}$  and  $a_{ij}$  the elements of  $\mathbf{E}$ , respectively  $\mathbf{A}$ .  $\mathbf{1}$  denotes a vector of 1's compatible with the dimensions of  $\mathbf{A}$ . Note that the first two constraints together restrict the solution to a simplex, *i.e.* they bound the  $l_1$ -norm of the solution. The constraints therefore already include the desired sparsity of the abundances (respectively, visible materials) in a pixel. By bounding the elements of  $\mathbf{E}$  also from above, we assert that the image intensities have been rescaled to surface reflectances (a material cannot reflect more than the incident energy). This can be achieved by rescaling the intensity values to the range  $[0 \dots 1]$ , assuming that there is at least one pure pixel in the image whose material is highly reflective in at least one spectral band.

### 4. Proposed Solution

To solve the super-resolution problem we need to estimate  $\mathbf{Z}$ , or equivalently  $\mathbf{E}$  and  $\mathbf{A}$ . From (2, 3, 4) we get the

following constrained least-squares problem:

$$\arg \min_{\mathbf{E}, \mathbf{A}} \|\mathbf{H} - \mathbf{EAS}\|_F^2 + \|\mathbf{M} - \mathbf{REA}\|_F^2 \quad (5a)$$

$$\text{subject to } 0 \leq e_{ij} \leq 1, \quad \forall i, j \quad (5b)$$

$$a_{ij} \geq 0, \quad \forall i, j \quad (5c)$$

$$\mathbf{1}^\top \mathbf{A} = \mathbf{1}^\top \quad (5d)$$

$$\|\mathbf{A}\|_0 \leq s \quad (5e)$$

with  $\|\cdot\|_F$  denoting the Frobenius norm, and  $\|\mathbf{A}\|_0$  the number of non-zero elements of  $\mathbf{A}$ . The constraints (5c, 5d) together restrict the abundances to the surface of a simplex spanned by the endmembers, and thus also act as a sparsity prior on the per-pixel abundances. The last constraint (5e) is optional, and serves to further increase sparsity, if desired.

Empirically, solving (5a) directly for  $\mathbf{E}$  and  $\mathbf{A}$  is difficult and rather unstable. The second term is strongly ill-posed w.r.t.  $\mathbf{E}$  due to the spectral degradation  $\mathbf{R}$ , in other words only  $b$  spectral channels do not contain sufficient information to separate  $p > b$  materials. Conversely, the first term is ill-posed w.r.t.  $\mathbf{A}$ , because the hyperspectral image, due to the downsampling  $\mathbf{S}$ , contains little information how to disentangle the abundance vector of a low-resolution pixel into contributions from its  $(N_m/N_h)$  constituent high-resolution pixels. We found it advantageous to split (5a) into a low-resolution ( $\mathbf{H}$ ) and a high-resolution ( $\mathbf{M}$ ) part and solve them by alternation.

The low-resolution step minimises the first term of (5a) subject to the constraints on  $\mathbf{E}$ ,

$$\arg \min_{\mathbf{E}} \|\mathbf{H} - \mathbf{E}\tilde{\mathbf{A}}\|_F^2 \quad (6)$$

$$\text{subject to } 0 \leq e_{ij} \leq 1, \quad \forall i, j.$$

*I.e.*, the endmembers of  $\mathbf{H}$  are updated for given low-resolution abundances  $\tilde{\mathbf{A}}$ . The latter are straight-forward to obtain from (preliminary estimates of) the high-resolution abundances  $\mathbf{A}$  by spatial downsampling, *c.f.* (2). The high-resolution step proceeds the opposite way and minimises the second term of (5a) under the constraints on  $\mathbf{A}$ ,

$$\arg \min_{\mathbf{A}} \|\mathbf{M} - \tilde{\mathbf{E}}\mathbf{A}\|_F^2$$

$$\text{subject to } a_{ij} > 0, \quad \forall i, j \quad (7)$$

$$\mathbf{1}^\top \mathbf{A} = \mathbf{1}^\top$$

$$\|\mathbf{A}\|_0 \leq s \quad (\text{optional})$$

This time the abundances at full resolution are updated for given endmembers  $\tilde{\mathbf{E}}$ , which are again just spectrally down-sampled version of the (preliminary) endmembers  $\mathbf{E}$  from the low-resolution step.

**Optimisation scheme.** Both parts of the alternation are constrained least-squares problems. Inspired by the

PALM (proximal alternating linearized minimisation) algorithm [8], we propose to use a projected gradient method for both parts. For (6) the following two steps are iterated for  $q = 1, 2, \dots$  until convergence:

$$\mathbf{U}^q = \mathbf{E}^{q-1} - \frac{1}{c}(\mathbf{E}^{q-1}\tilde{\mathbf{A}} - \mathbf{H})\tilde{\mathbf{A}}^\top \quad (8a)$$

$$\mathbf{E}^q = \text{prox}_{\mathbf{E}}(\mathbf{U}^q) \quad (8b)$$

with  $c = \gamma_1 \|\tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top\|_F$  a non-zero constant, and  $\text{prox}_{\mathbf{E}}$  a proximal operator that projects onto the constraints of (6). What makes the algorithm attractive is that  $\text{prox}_{\mathbf{E}}$  is computationally very cheap: it amounts to truncating the entries of  $\mathbf{V}$  to 0 from below and to 1 from above.

Likewise, (7) is minimised by iterating the following steps until convergence:

$$\mathbf{V}^q = \mathbf{A}^{q-1} - \frac{1}{d}\tilde{\mathbf{E}}^\top(\tilde{\mathbf{E}}\mathbf{A}^{q-1} - \mathbf{M}) \quad (9a)$$

$$\mathbf{A}^q = \text{prox}_{\mathbf{A}}(\mathbf{V}^q) \quad (9b)$$

with  $d = \gamma_2 \|\tilde{\mathbf{E}}\tilde{\mathbf{E}}^\top\|_F$  again a non-zero constant and  $\text{prox}_{\mathbf{A}}$  a proximal operator that projects onto the constraints of (7). Again the proximal operator for the simplex projection is computationally efficient, see [10]. The complete optimisation scheme is given in Alg. 1.

Since (5a) is highly non-convex, we need good initial values to start the local optimisation. We choose SISAL [4] to initialise the endmembers. SISAL robustly fits a minimum-volume simplex to the response vectors (the columns) of  $\mathbf{H}$  with a sequence of augmented Lagrangian optimisations, and returns the vertices of the simplex as endmembers. With the initial endmembers  $\mathbf{E}^{(0)}$  we then use SUnSAL [5] to get initial abundances. SUnSAL includes the constraints (5c, 5d) and solves a constrained least-squares problem for  $\tilde{\mathbf{A}}^{(0)}$ , via alternating direction method of multipliers (ADMM). Finally we compute  $\mathbf{A}^{(0)}$  by upsampling  $\tilde{\mathbf{A}}^{(0)}$ . Formally, this can be seen as applying the pseudo-inverse of the downsampling operator,

---

**Algorithm 1** Solution of minimisation problem (5a).

---

**Require:**  $\mathbf{H}, \mathbf{M}, \mathbf{S}, \mathbf{R}$

Initialize  $\mathbf{E}^{(0)}$  with SISAL and  $\tilde{\mathbf{A}}^{(0)}$  with SUnSAL

Initialize  $\mathbf{A}^{(0)}$  by upsampling  $\tilde{\mathbf{A}}^{(0)}$

$k \leftarrow 0$

**while** not converged **do**

$k \leftarrow k + 1$

// low-resolution step:

$\tilde{\mathbf{A}} \leftarrow \mathbf{A}^{(k-1)}\mathbf{S}$  ; Estimate  $\mathbf{E}^{(k)}$  with (8a, 8b)

// high-resolution step:

$\tilde{\mathbf{E}} \leftarrow \mathbf{R}\mathbf{E}^{(k)}$  ; Estimate  $\mathbf{A}^{(k)}$  with (9a, 9b)

**end while**

**return**  $\mathbf{Z} = \mathbf{E}^{(k)}\mathbf{A}^{(k)}$

---

$\mathbf{A}^{(0)} = \tilde{\mathbf{A}}^{(0)}(\mathbf{S}^\top \mathbf{S})^{-1} \mathbf{S}^\top$ . Empirically, it is beneficial to smooth the initial abundance image  $\mathbf{A}^{(0)}$  a bit to avoid blocking artifacts (although these do not greatly impact the quantitative results).

In our experiments we set  $\gamma_1 = \gamma_2 = 1.01$ . These values only affect the speed of convergence. In our experience the optimisation exhibits monotonic convergence.

## 5. Experiments

**Datasets.** To evaluate our method we use two publicly available hyperspectral databases. The first database, called CAVE [28], includes 32 indoor images showing *e.g.* paintings, toys, food, *etc.*, captured under controlled illumination. The dimensions of the images are  $512 \times 512$  pixels, with 31 spectral bands, each 10 nm wide, covering the visible spectrum from 400 to 700 nm. The second database, called Harvard [9], has 50 indoor and outdoor images recorded under daylight illumination, and 27 images under artificial or mixed illumination. The spatial resolution of these images is  $1392 \times 1040$  pixels, with 31 spectral bands of width 10 nm, ranging from 420 to 720 nm. We use only the top left  $1024 \times 1024$  pixels (to avoid fractional coverage of the low-resolution pixels, see below). The original images in the databases serve as ground truth. To obtain low-resolution hyperspectral images  $\mathbf{H}$  we downsample them by a factor of 32, to create  $16 \times 16$  images. The same procedure was used in [16, 27, 1]. The downsampling is done by simple averaging over  $32 \times 32$  pixel blocks. The conventional images  $\mathbf{M}$  were created by integrating over the original spectral channels, using the spectral response  $\mathbf{R}$  of a typical digital camera<sup>2</sup>. The area under the response curve is normalised to 1 to ensure all three colour channels have the same intensity range.

**Implementation details.** We run our method with the maximum number of endmembers set to  $p = 10$ , which is sufficient for both datasets. The inner loops of the two optimisation steps (8a), (9a) are run until the update falls below 1%, which typically takes  $\approx 10$  iterations in the early stages and drops to  $\approx 2$  iterations as the alternation proceeds. The outer loop over the two steps is iterated until the overall cost (5a) changes less than 0.01%, or for at most 1500 iterations.

Our current implementation in Matlab 8.4 has not been optimised for speed. Computation times depend on the image size and the number of iterations, as well as the sparsity parameter  $s$ , if used. For a  $512 \times 512$  pixels image with 31 channels it takes  $\approx 9$  minutes, on a single Intel Xeon E5 3.2 Ghz CPU.

**Baselines.** As baselines to compare against, we use five state-of-the-art methods [1, 2, 21, 27, 29], in the authors'

<sup>2</sup>Nikon D700, [www.maxmax.com/spectral\\_response.htm](http://www.maxmax.com/spectral_response.htm)

Method	CAVE database [28]					
	RMSE			SAM		
	Aver.	Med.	1st	Aver.	Med.	1st
Akhtar' 14 [1]	5.1	5.0	0	12.5	12.1	0
Akhtar' 15 [2]	5.0	4.6	0	12.0	12.2	0
HySure [21]	4.9	4.2	0	21.0	21.9	0
Wycoff [27]	4.5	4.2	1	18.8	19.0	0
Yokoya [29]	3.5	3.2	5	6.2	6.2	11
Ours	<b>3.0</b>	<b>2.6</b>	<b>26</b>	<b>5.8</b>	<b>6.0</b>	<b>21</b>

Table 1. Results for the CAVE database.

original implementations. The baselines were chosen because they reported the best results in the literature, and the code for all five was made available by the authors. All methods were run with the same spectral response  $\mathbf{R}$ , and tuned for best performance. For [1] we follow the original paper and use a dictionary with  $k = 75$  atoms.<sup>3</sup> We tested larger numbers, too, but this did not improve the performance. For [21] we use only the fusion part and deactivate the additional estimation of the point spread function (which is perfectly known, so re-estimating it would give the method a disadvantage).

**Error metrics.** As primary error metric, we use the root mean square error of the estimated high-resolution hyperspectral image  $\hat{\mathbf{Z}}$  w.r.t. the ground truth image  $\mathbf{Z}$ , on an 8-bit intensity range  $[0 \dots 255]$ .

$$\text{RMSE} = \sqrt{\frac{1}{BN_m} \sum \sum (\hat{z}_{ij} - z_{ij})^2} = \sqrt{\frac{\|\hat{\mathbf{Z}} - \mathbf{Z}\|_F^2}{BN_m}} \quad (10)$$

Additionally, we also use the the spectral angle mapper (SAM, [30]), which is defined as the angle in  $\mathbb{R}^B$  between the estimated pixel  $\hat{\mathbf{z}}_j$  and the ground truth pixel  $\mathbf{z}_j$ , averaged over the whole image.

$$\text{SAM} = \frac{1}{N_m} \sum \arccos \frac{\hat{\mathbf{z}}_j^\top \mathbf{z}_j}{\|\hat{\mathbf{z}}_j\|_2 \|\mathbf{z}_j\|_2} \quad (11)$$

where  $\|\cdot\|_2$  is the  $l_2$  vector norm. The SAM is given in degrees.

We also report, for both metrics, for how many images in a database each method returned the best result. This number gives an intuition whether a method is consistently better than another one across different images.

<sup>3</sup>The results for [1, 2] differ from the originally published ones. According to a personal communication with the authors, the quantity labelled "RMSE" in those papers is in fact the spectral norm of the residual matrix (rather than its Frobenius norm), and thus not comparable.

Method	Harvard database [9]					
	RMSE			SAM		
	Aver.	Med.	1st	Aver.	Med.	1st
Akhtar'14 [1]	2.7	2.2	0	4.0	3.7	0
Akhtar'15 [2]	2.4	1.8	4	4.4	3.8	0
HySure [21]	2.4	1.9	0	5.0	4.2	0
Wycoff [27]	2.4	1.8	2	5.1	4.4	1
Yokoya [29]	<b>1.7</b>	<b>1.5</b>	22	<b>2.9</b>	<b>2.7</b>	35
Ours	<b>1.7</b>	<b>1.5</b>	<b>49</b>	<b>2.9</b>	<b>2.7</b>	<b>41</b>

Table 2. Results for the Harvard database.

**Experimental results.** Tables 1 and 2 show the average and median RMSE and SAM values across the two datasets, as well as the number of images that each method reconstructed best. A complete table of results for the individual images is given in the supplementary material.

The proposed joint unmixing approach achieves significantly lower RMSE than all five baselines for the CAVE database. The method reduces the error by  $\approx 15\%$  compared to the best baseline and more than 30% against all others. It also returns the best result for the overwhelming majority of images. Also in terms of SAM, our method is clearly best across all three error statistics, although [29] comes relatively close. Compared to all other baselines, the improvement in SAM is even greater, in the range of 30-50%. These results suggest that our method and [29] are particularly good at reconstructing the spectral distribution of the intensities, and perform evenly over all spectral bands.<sup>4</sup> The Harvard database is significantly easier, consequently the reconstruction errors are a factor  $\approx 2$  lower, for all methods. Again coupled factorisation approaches seem to have the upper hand, with our method narrowly beating [29].

Interestingly, among the baselines [21, 27] perform clearly worse than the others in terms of SAM, but somewhat better in terms of RMSE. Apparently these methods do well in reconstructing the relative proportions of some bands, whereas for some other bands the errors are distributed more unevenly across the spectrum. We note that [27] has no mechanism to prevent all coefficients for a pixel becoming 0, and thus sometimes produces pixels with intensity 0 in all channels. These pixels were excluded from the computation of SAM (which is not defined for vectors of length 0). Similarly, [2] constrains neither the basis nor the coefficients to be  $\geq 0$ , and produces a few pixels with negative intensities in most images. These pixels were truncated to 0 for the evaluation.

<sup>4</sup>In the literature, higher errors, and also unrealistically long runtimes, are reported for [29]. Those numbers, first published in [27] and copied in [1, 2], appear to be wrong. We generated the numbers in Tab. 1, 2 with N. Yokoya's original code, and suggest to use them for future reference.

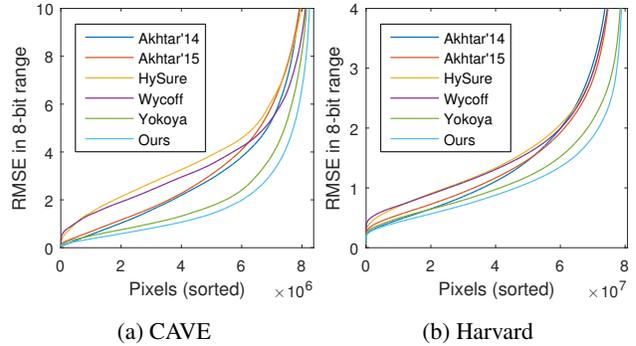


Figure 2. Per-pixel RMSE over all database images, sorted by magnitude.

Generally, the three methods that explicitly use the downsampling operator  $\mathbf{S}$ , [27, 29] and ours, seem to outperform methods that do not relate the low- and high-resolution abundances through  $\mathbf{S}$ . In our view this is not a big limitation, since it takes little effort to calibrate the point spread function, or even estimate it from the images.

To complement the tabulated results, we also visualize the distribution of the per-pixel residuals across the complete datasets. In Figure 2 all the pixels of each dataset are sorted according to their per-pixel RMSE, computed over the 31 spectral channels. The graph shows that with our method a significantly larger number of pixels have very small reconstruction errors below 1, respectively 2 gray-values, and our results also exhibit the smallest fraction of grossly wrong pixels with errors above 4 gray values. Example images and estimated abundance maps are shown in Figs. 3 and 4. Examples of hyperspectral channels reconstructed by our method are shown in Fig. 5.

**Effect of the sparsity term.** We also tested the effect of the optional sparsity term (5e). Although the simplex constraints already favours sparse solutions, it may in some situations be desired to suppress small abundance values and enforce stronger sparsity. Note, sparser solutions may have more realistic endmembers, but they will in general have higher reconstruction error, because the solution is more constrained. Using only the simplex constraints, each pixel in the CAVE database has on average 6.5 active endmembers. Accordingly, we test values  $s \in \{5N_m, 4N_m, 3N_m, 2N_m\}$  to obtain progressively sparser solutions. Table 3 shows the corresponding results. As expected, the error initially increases slowly, indicating that a faithful reconstruction of the image is possible even with only  $4N_m$  non-zero abundances. Below this value, the error starts to increase more rapidly. Note, though, that even with 3 active endmembers on average, our reconstruction errors are as low as those of the best baselines, see Tab. 1.

CAVE database [28]					
$s/N_m$	—	5	4	3	2
RMSE	3.00	3.05	3.16	3.51	4.86
SAM	5.81	6.03	6.39	7.11	8.24

Table 3. Effect of the sparsity parameter  $s$ . The first row is the selected sparsity level (average number of non-zero endmembers per pixel). The values in the first column are without the additional sparsity term.

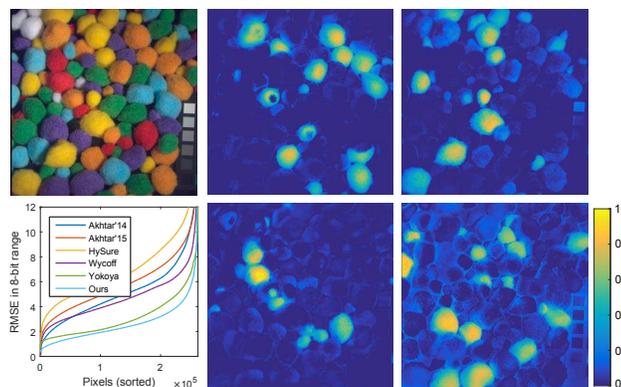


Figure 3. *Top right*: Example image from the CAVE database. *Bottom left*: per-pixel residuals of different methods, sorted by magnitude. Our method produces significantly fewer outliers with high reconstruction error. *Centre and right*: Abundances nicely capture differently coloured pompons.

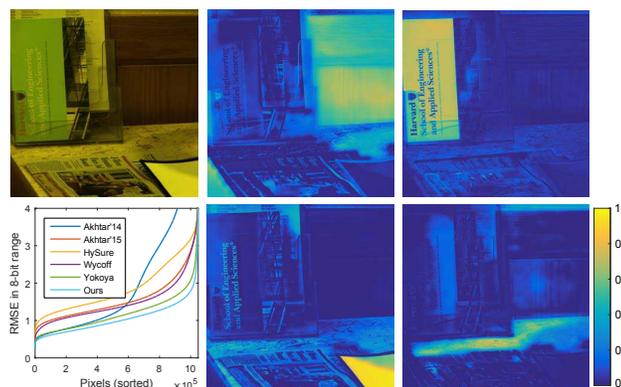


Figure 4. Example image from the Harvard database. *Bottom left*: Per-pixel RMSE sorted by magnitude. *Center and right*: Abundances capture different materials.

## 6. Discussion

**Super-resolution vs. unmixing.** The evaluation shows that we are able to generate a super-resolution image which is upsampled by a factor of 32 in the spatial domain, respec-

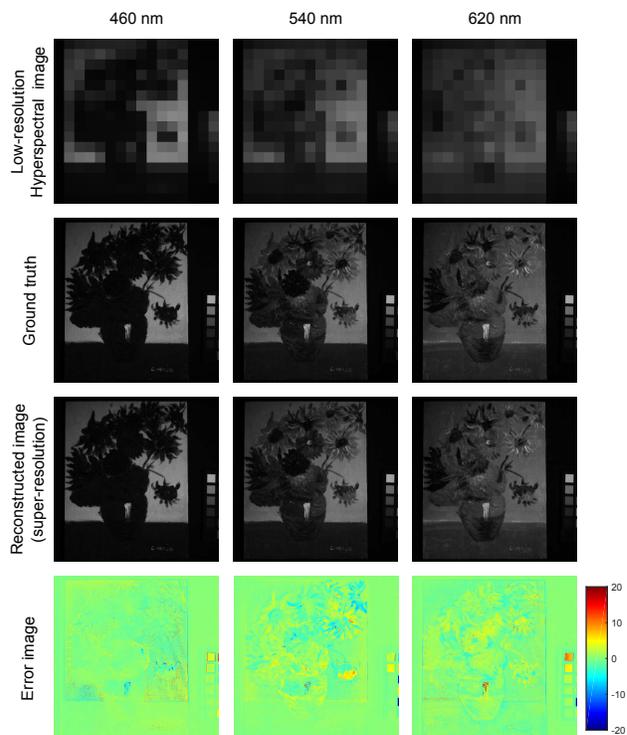


Figure 5. Spectral images of a painting from the CAVE database. Depicted are the spectral bands at 460, 540 and 620 nm. *Top row*: Low resolution input images. *Second row*: Ground truth. *Third row*: Reconstructed images from our method. *Bottom row*: Reconstruction errors in 8-bit range.

tively a factor of 10 in the spectral domain, compared to the two inputs. While this gives some indication that the spectral unmixing in our method is successful, we cannot evaluate whether it is actually correct, since there is no ground truth available.

The proposed method in its basic form has only one user parameter, namely the number  $p$  of endmembers. If additional sparsity is desired beyond the simplex constraint, then this introduces a second parameter, the number of non-zero abundances  $s$ . If high-quality super-resolution is the goal, then it is generally better to take a conservative approach and set  $p$  larger than the number of actual materials in the scene. This will cause multiple endmembers for some materials, e.g. to cover brightness differences due to shading and shadows, but will nevertheless deliver good super-resolution. In fact, it can even be helpful to cover non-linear effects with additional endmembers. While our work once again confirms that the LMM is sufficient for many hyperspectral imaging problems and non-linear effects by and large fade in the noise, there are some exceptions. In particular, specular reflections are almost unavoidable, but can

easily be covered by a single “endmember” that has very high reflectance across all spectral channels.

We found that the single most important ingredient of our method, and the main reason for its excellent performance, is the simplex constraint (5c, 5d). The two requirements that the endmembers must be positive *and* sum to 1 together greatly stabilise the model, and have not been used in this form in any competing approach (although several authors have used one of the two constraints, see Sec. 2). Increased sparsity through a stricter  $s$  can reduce over-fitting and lead to a physically more correct unmixing by suppressing overly small abundances, but will usually not improve the super-resolution, see Tab. 3. Note that, in contrast to some other methods for super-resolution or unmixing, our algorithm will always keep at least one endmember with non-zero abundance at every pixel, thanks to the sum-to-one constraint (assuming  $s > N_m$ , otherwise the solution space is empty).

**Datasets.** The two datasets have some peculiarities. In both datasets images were taken with the help of tunable filters, and thus different channels were not acquired simultaneously. For the CAVE database the sensor used a fixed focal length, which results in out-of-focus blur at the extremes of the spectral range. In the super-resolution images there is no such blurring, since they get their high-frequency content mostly from the (sharp) high-resolution conventional image. Figure 6 shows an example where the super-resolution reconstruction for the 410 nm channel is visually clearly better than the original image. This problem leads to a small bias in the error measures (for any method).

In the case of the Harvard database, some objects have moved during acquisition, especially in the outdoor images (*e.g.* trees waving in the wind). Since most models assume pixel-accurate registration between the inputs, the super-resolution image will have artefacts, and as a consequence higher RMSE. Except for these cases, the Harvard dataset is noticeably easier. All methods achieve significantly lower errors than on CAVE, which has a larger number and variability of materials per image, so that super-resolution is more challenging.

Overall, it seems that further research into hyperspectral imaging will require new datasets, which on one hand are larger and more challenging (RMSE is already down to  $\approx 1\%$  of the intensity range), and on the other hand avoid imaging problems, especially wavelength-dependent blur. New hyperspectral sensors (*e.g.* `www2.imec.be`) may be a good tool to acquire such data.

## 7. Conclusions

We have proposed a new method for hyperspectral super-resolution. The basic idea is to jointly solve the spectral

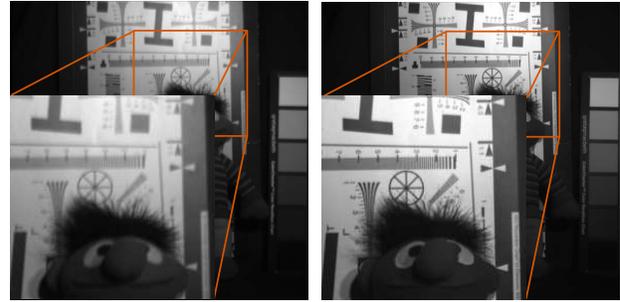


Figure 6. *Left*: Ground truth image for the 410 nm channel. *Right*: Reconstructed image from our method. The ground truth exhibits out-of-focus blur, whereas the reconstructed image does not.

unmixing problem for both input images. Linking super-resolution to unmixing allows one to impose a number of constraints, which are due to the elementary physics of the imaging process. Under the linear mixing model the proposed approach boils down to two coupled, constrained least-squares problems, which can be solved reliably with a projected gradient scheme. In experiments on two public datasets the method has shown excellent performance.

In future work, we aim to extend the method in several directions. To deal with larger scenes, *e.g.* in remote sensing imagery, it may be useful to employ a locally varying set of endmembers, so as to adapt to changes of the environment and/or lighting effects without unnecessarily inflating the basis. Remote sensing images may also require an extension of the strict mixing model, in order to deal with atmospheric influences and possibly even non-linear effects, *e.g.* volume scattering in the vegetation. A different direction would be to include mixed sets of multiple hyperspectral and conventional images of the same scene, to further improve the estimation, or even move towards additional reflectance parameters beyond a diffuse spectral albedo.

## References

- [1] N. Akhtar, F. Shafait, and A. Mian. Sparse spatio-spectral representation for hyperspectral image super-resolution. In *Computer Vision–ECCV 2014*, pages 63–78. Springer, 2014.
- [2] N. Akhtar, F. Shafait, and A. Mian. Bayesian sparse representation for hyperspectral image super resolution. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 3631–3640, 2015.
- [3] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3012–3021, 2007.
- [4] J. M. Bioucas-Dias. A variable splitting augmented lagrangian approach to linear spectral unmixing. In *First Workshop on Hyperspectral Image and Signal Processing: Evolution*

- tion in Remote Sensing (WHISPERS'09), pages 1–4. IEEE, 2009.
- [5] J. M. Bioucas-Dias and J. M. Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8):2435–2445, 2008.
- [6] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE, Geoscience and Remote Sensing Magazine*, 1(2):6–36, 2013.
- [7] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):354–379, 2012.
- [8] J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, 2014.
- [9] A. Chakrabarti and T. Zickler. Statistics of Real-World Hyperspectral Images. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 193–200, 2011.
- [10] L. Condat. Fast projection onto the simplex and the  $l_1$  ball, 2014. [Online] <http://hal.univ-grenoble-alpes.fr/hal-01056171/>.
- [11] Y. Gu, Y. Zhang, and J. Zhang. Integration of spatial-spectral information for resolution enhancement in hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5):1347–1358, 2008.
- [12] R. C. Hardie, M. T. Eismann, and G. L. Wilson. Map estimation for hyperspectral image resolution enhancement using an auxiliary sensor. *IEEE Transactions on Image Processing*, 13(9):1174–1184, 2004.
- [13] B. Huang, H. Song, H. Cui, J. Peng, and Z. Xu. Spatial and spectral image fusion using sparse matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 52(3):1693–1704, 2014.
- [14] F. H. Imai and R. S. Berns. High-resolution multi-spectral image archives: a hybrid approach. In *6th Color and Imaging Conference*, volume 1998, pages 224–227. Society for Imaging Science and Technology, 1998.
- [15] T. Kasetkasem, M. K. Arora, and P. K. Varshney. Super-resolution land cover mapping using a markov random field based approach. *Remote Sensing of Environment*, 96(3):302–314, 2005.
- [16] R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi. High-resolution hyperspectral imaging via matrix factorization. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 2329–2336. IEEE, 2011.
- [17] N. Keshava and J. F. Mustard. Spectral unmixing. *IEEE Signal Processing Magazine*, 19(1):44–57, 2002.
- [18] S. J. Kim, F. Deng, and M. S. Brown. Visual enhancement of old documents with hyperspectral imaging. *Pattern Recognition*, 44(7):1461–1469, 2011.
- [19] Z. Pan, G. Healey, M. Prasad, and B. Tromberg. Face recognition in hyperspectral images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1552–1560, 2003.
- [20] C. Pohl and J. L. V. Genderen. Review article multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing*, 19(5):823–854, 1998.
- [21] M. Simoes, J. Bioucas-Dias, L. Almeida, and J. Chanussot. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6):3373–3388, 2015.
- [22] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition*, 43(7):2367–2379, 2010.
- [23] H. Van Nguyen, A. Banerjee, and R. Chellappa. Tracking via object reflectance using a hyperspectral video camera. In *IEEE Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 44–51. IEEE, 2010.
- [24] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li. A comparative analysis of image fusion methods. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1391–1402, 2005.
- [25] Q. Wei, N. Dobigeon, and J.-Y. Tourneret. Bayesian fusion of hyperspectral and multispectral images. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3176–3180. IEEE, 2014.
- [26] D. Wu and D.-W. Sun. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A reviewpart i: Fundamentals. *Innovative Food Science & Emerging Technologies*, 19:1–14, 2013.
- [27] E. Wycoff, T.-H. Chan, K. Jia, W.-K. Ma, and Y. Ma. A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1409–1413. IEEE, 2013.
- [28] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. Technical report, Nov. 2008.
- [29] N. Yokoya, T. Yairi, and A. Iwasaki. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):528–537, 2012.
- [30] R. H. Yuhas, A. F. Goetz, and J. W. Boardman. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In *Summaries of the Third Annual JPL Airborne Geoscience Workshop*, pages 147–149, 1992.
- [31] H. Zhang, L. Zhang, and H. Shen. A super-resolution reconstruction algorithm for hyperspectral images. *Signal Processing*, 92(9):2082–2096, 2012.