ADVANCES IN HYPERSPECTRAL AND MULTISPECTRAL IMAGE FUSION AND SPECTRAL UNMIXING

C. Lanaras,[*]   **E. Baltsavias**,   **K. Schindler**

Institute of Geodesy and Photogrammetry, ETH Zürich, Switzerland
{ lanaras | manos | schindler }@geod.baug.ethz.ch

KEY WORDS: hyperspectral, image fusion, super-resolution, spectral unmixing, endmembers, processing techniques

ABSTRACT:

In this work, we jointly process high spectral and high geometric resolution images and exploit their synergies to (a) generate a fused image of high spectral *and* geometric resolution; and (b) improve (linear) spectral unmixing of hyperspectral endmembers at subpixel level w.r.t. the pixel size of the hyperspectral image. We assume that the two images are radiometrically corrected and geometrically co-registered. The scientific contributions of this work are (a) a simultaneous approach to image fusion and hyperspectral unmixing, (b) enforcing several physically plausible constraints during unmixing that are all well-known, but typically not used in combination, and (c) the use of efficient, state-of-the-art mathematical optimization tools to implement the processing. The results of our joint fusion and unmixing has the potential to enable more accurate and detailed semantic interpretation of objects and their properties in hyperspectral and multispectral images, with applications in environmental mapping, monitoring and change detection. In our experiments, the proposed method always improves the fusion compared to competing methods, reducing RMSE between 4% and 53%.

1. INTRODUCTION

The aim of this paper is the integrated processing of (a) high spectral resolution images (HSRI) with low geometric resolution, and (b) high geometric resolution images (HGRI) with low spectral resolution. From a pair of one HSRI and one HGRI, the integrated processing (a) generates a fused image with both high spectral and geometric resolution (Figure 1). This can be thought of as a sort of hyperspectral pan-sharpening, improved through the use of spectral unmixing constraints; and (b) a better spectral unmixing of the HSRI, through integration of geometric information from the HGRI.

Hyperspectral imaging, often also called imaging spectroscopy, delivers images with many (up to several hundreds) contiguous spectral bands of narrow bandwidth, typically in the order of a few nm. These bands may cover different spectral ranges, e.g. 0.4-$2.5\mu m$, 0.4-$1.0\mu m$ or 0.4-$0.7\mu m$. The geometric resolution (or ground sampling distance GSD) of HSRI is low, on the one hand due to physical constraints (i.e. with a narrow bandwidth the signal is weak, so GSD has to be increased to yield an acceptable SNR), and on the other hand due to the possible constraints of the platform (satellites) for storing and/or transmitting a large amount of data, making it necessary to sacrifice geometric resolution in favour of spectral resolution. The HGRI are typically multispectral images with often only 3 to 4 bands in the RGB and NIR bands, with a wide bandwidth of several tens of nm. Theoretically, HGRI could also be panchromatic images, although we currently do not handle this case. HSRI are increasingly available, especially from airborne and satellite platforms, but lately also UAVs and terrestrial photography (for a list of some sensors see e.g. Varshney and Arora, 2004). HGRI are widely acquired from all kind of platforms. These two types of images are complementary to some degree, but have rarely been processed jointly. Merging them into a single, enhanced image product is, especially in the computer vision literature, sometimes termed "hyperspectral image fusion" or "hyperspectral super-resolution". The images can be acquired (quasi) simultaneously from the same platform
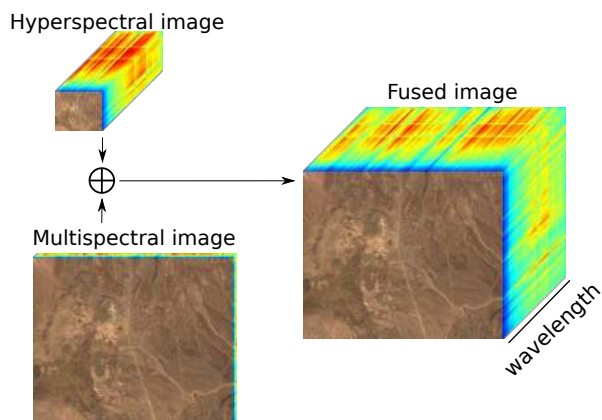

Figure 1. Hyperspectral images have high spectral resolution but low geometric resolution, whereas the opposite is true for multispectral images. In this paper, we aim at fusing the two types of imagery.

(e.g. sensors Hyperion and ALI on the EO-1 satellite), or with a time difference. The latter case is more frequent – HSRI and HGRI sensors are rarely used simultaneously – but also more difficult, because multitemporal differences (e.g. different illumination, atmospheric conditions, scene changes) introduce errors. We currently assume that the images have already been radiometrically corrected, that the bands of each image and the two images are geometrically accurately co-registered, and that a dimensionality reduction of the hyperspectral bands has been performed, if needed.

The proposed integrated processing shall serve as basis for more accurate and detailed semantic information extraction. HSRI can distinguish finer spectral differences among objects. Thus, not only more objects can be localized and classified, e.g. for land-cover mapping, but also biophysical/biochemical properties of objects can be better estimated (for a list of some applications see e.g. Varshney and Arora, 2004). Since HSRI have a large GSD, mixed pixels occur more often, especially in settlement areas.

---

[*]Corresponding author

The task of unmixing is to determine for each pixel the "endmembers" (so called "pure" material spectra) and the proportional contribution of each endmember within the pixel (called "fractional abundance", or simply "abundance"). The abundances andendmembers are subject to several physical constraints, like that the abundances for all endmembers in a pixel must be non-negative and sum to 1, while the number of endmembers present within each pixel, is rather low. To the best of our knowledge, though, there is not yet a fusion method which simultaneously uses all available constraints.

Regarding unmixing, there are two main classes of models, the linear mixing model (LMM) and different nonlinear ones (see e.g. Keshava and Mustard, 2002). In the first case, one assumes only direct reflections, and abundances represent areas of endmembers within a pixel. In the non-linear case, multiple reflections are taken into account and multiple endmembers can be overlayed in the radiometric signature of a single surface location (see Heylen et al., 2014, for a nonlinear unmixing review). Most research uses the simpler LMM, which we also do here. The reason to include HGRI in the unmixing lies in the higher geometric resolution. Geometric information, e.g. homogeneous segments, edges and texture boundaries at a resolution higher than the GSD of the HSRI can help to achieve a more accurate unmixing, and also to spatially localize abundances: rather than only estimating the contribution (%) of one endmember within a pixel, one can determine the corresponding regions (i.e. the pixels at the higher HGRI resoution).

This work makes the following contributions: At the conceptual level we process HSRI and HGRI together to simultaneously perform image fusion and spectral unmixing. This coupled estimation lets the two processes support each other and leads to better results. At the level of modelling, one needs to adequately couple the two processes, and we do this by imposing multiple physically motivated constraints on the endmembers and abundances in both input images. Thus, we treat LMM as a coupled, constrained matrix factorization of the two input images into endmembers and their abundances. On the technical level, we employ efficient state-of-the-art optimization algorithms to tackle the resulting constrained least-squares problem (Bolte et al., 2014, Condat, 2014). Experimental results on several different image pairs show a consistent improvement compared to several other fusion methods.

2. RELATED WORK

Here, due to lack of space we focus on research on hyperspectral image processing. In recent years, hyperspectral imaging is advancing rapidly (Plaza et al., 2009, Ben-Dor et al., 2013). Both geometric and spectral resolution have improved, which opens up new applications in different fields, including geology and prospecting, agriculture, ecosystems, earth sciences, etc. The fine spectral resolution allows for fine material characterization within one pixel, while the low geometric resolution is a limiting factor. According to Bioucas-Dias et al. (2013) hyperspectral remote sensing data analysis can be grouped into five main areas of activity: hyperspectral unmixing, data fusion, classification, target detection and estimation of land physical parameters. The first two parts are most relevant for this research. The case of the LMM (for a recent review see Bioucas-Dias et al., 2012) is the best studied one, and sufficient for many purposes. To solve it, the first step is to identify the number of endmembers as well as their spectra. The so-called endmember extraction algorithms can be categorized as follows.

Geometrical approaches exploit the fact that in a linear mixture the observed spectral responses form a simplex. Among them,

pure pixel algorithms identify a set of "pure" pixels in the data and assume that these are the vertices of the simplex. Well-known examples include vertex component analysis (VCA) (Nascimento and Bioucas Dias, 2005), AVMAX and SVMAX (Chan et al., 2011). *Minimum volume algorithms* on the other hand do not assume that pure pixels exist, but rather that some data points exist on the facets of a simplex. Examples include SISAL (Bioucas-Dias, 2009), MVC-NMF (Miao and Qi, 2007) and the sparsity-promoting SPICE (Zare and Gader, 2007). *Statistical approaches* are not limited to the surface of the simplex (e.g. DECA, Nascimento and Bioucas-Dias, 2012), and can deal better with highly mixed scenarios, at the cost of increased complexity. Finally (sparse) *regression approaches* formulate unmixing as a sparse linear regression problem, assuming that the "pure" spectral signatures are given in advance. With this assumption, the observations are mixtures of a subset of given signatures available in a "spectral library".

The inversion step consists of solving for the fractional abundances under additional constraints such as non-negativity and sum-to-one. This can be done with direct least squares estimation (Heinz and Chang, 2001) or basis pursuit, including variants regularized with total variation (Iordache et al., 2012).

Hyperspectral data fusion denotes the problem of combining a HSRI with another image of higher geometric resolution, usually a multispectral one. Most methods rely on some form of matrix factorization. Yokoya et al. (2012) first extract endmembers with standard VCA and then apply a Non-negative Matrix Factorization (NMF) to jointly estimate the abundances for both a HSRI and a HGRI in an iterative manner. Akhtar et al. (2014) also extract reflectance spectra to form a spectral basis in advance, and with the basis fixed use a non-negative OMP (Orthogonal Matching Pursuit) to learn a sparse coding for the image. A related approach (Kawakami et al., 2011) estimates the endmembers by sparse dictionary learning and further enforces the multispectral abundances to be sparse. However, results may not be physically plausible, since non-negativity is not considered. Huang et al. (2014) learn the spectral basis with SVD (Singular Value Decomposition) and solve for the high spectral resolution abundances via OMP. Wycoff et al. (2013) introduce a non-negative sparsity-promoting framework, which is solved with the ADMM (Alternating Direction Method of Multipliers) solver. Simoes et al. (2015) formulated a convex subspace-based regularization problem, which can also be solved with ADMM, and estimate also the spectral response and spatial blur from the data. However, they used a fixed subspace for the spectral vectors, not considering any physical interpretation. Bayesian approaches (Hardie et al., 2004, Wei et al., 2014) additionally impose priors on the distribution of image intensities and perform Maximum A Posteriori (MAP) inference. Kasetkasem et al. (2005) also add a Markov Random Field (MRF) prior to model spatial correlations. It has also been suggested to learn a prior for the local structure of the upsampled images from offline training data (Gu et al., 2008). Such blind deblurring will however work at most for moderate upsampling.

3. PROBLEM FORMULATION

The goal of fusion is to combine information coming from a hyperspectral image (HSRI) $\bar{\mathbf{H}} \in \mathbb{R}^{w \times h \times B}$ and a multispectral image (HGRI) $\bar{\mathbf{M}} \in \mathbb{R}^{W \times H \times b}$. The HSRI has high spectral resolution, with B spectral bands, but low geometric resolution with w, h, image width and height respectively. The HGRI has high geometric resolution, with W, H, image width and height respectively, but a low spectral resolution b. We want to estimate a fused image $\bar{\mathbf{Z}} \in \mathbb{R}^{W \times H \times B}$ that has both high geometric and

high spectral resolution. For this fused image we expect to have the spectral resolution of the HSRI (a lot larger than the HGRI, $B \gg b$) and the geometric resolution of the HGRI (a lot larger than the HSRI, $W \gg w$ and $H \gg h$).

To simplify the notation, we treat each image ("data cube") as a matrix, with each row holding all pixels of an image in a certain spectral band, i.e. $N_h = wh$ and $N_m = WH$ pixels per row, while each matrix column shows all spectral values for a given pixel. Accordingly, the images are written $\mathbf{Z} \in \mathbb{R}^{B \times N_m}$, $\mathbf{H} \in \mathbb{R}^{B \times N_h}$ and $\mathbf{M} \in \mathbb{R}^{b \times N_m}$.

Following the LMM (Keshava and Mustard, 2002, Bioucas-Dias et al., 2012), mixing accurs additively within each pixel and the spectral response $\mathbf{z} \in \mathbb{R}^B$ of a given pixel of $\mathbf{Z}$ is approximated as

$$\mathbf{z} = \sum_{j=1}^{p} \mathbf{e}_j a_j \qquad (1)$$

where $\mathbf{e}_j \in \mathbb{R}^B$ denotes the spectral measurement (e.g. reflectance) of the $j \in \{1, \ldots, p\}$ endmember, a_j denotes the *fractional abundance* of the respective endmember j, and p is the number of endmembers used to explain this pixel. Eq. (1) can be rewritten in matrix form:

$$\mathbf{Z} = \mathbf{EA} \qquad (2)$$

where $\mathbf{E} \equiv [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p]$, $\mathbf{E} \in \mathbb{R}^{B \times p}$, $\mathbf{A} \equiv [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{N_m}]$, $\mathbf{A} \in \mathbb{R}^{p \times N_m}$ and $\mathbf{a}_i \in \mathbb{R}^p$ are the fractional abundances for every pixel $i = \{1, 2, \ldots, N_m\}$. By using this expression, we are able to represent the image $\mathbf{Z}$ in a basis spanned by the endmembers, with the abundances as coefficients. Given that $p < B$, we are able to describe $\mathbf{Z}$ in a lower dimensional space, thus achieving a dimensionality reduction. The assumption that $p < B$ and consequently that the data "live" in a lower dimensional space holds in most the hyperspectral images and thus $rank\{\mathbf{Z}\} \leq p$.

3.1 Constraints

The basic idea of the proposed fusion and unmixing is to use the fact that in (1) the endmembers and abundances have a physical meaning and they can retain this meaning only under the following constraints:

$$
\begin{aligned}
a_{ij} &\geq 0 \ \forall i,j & \text{(non-negative abundance)} \\
\mathbf{1}^\top \mathbf{A} &= \mathbf{1}^\top & \text{(abundances sum to 1)} \\
0 &\leq e_{ij} \leq 1 \ \forall i,j & \text{(non-negative, bounded reflectance)}
\end{aligned} \qquad (3)
$$

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 $\mathbf{A}$.

The first two constraints restrict the solution to a simplex and promote sparsity. The third constraint bounds the endmembers from above and implies that a material cannot reflect more than the incident radiation. Even if the image values do not represent reflectance, we can rescale the values in the range $[0 \ldots 1]$, assuming that there are some materials in the scene that are higly reflective in a part of the spectrum.

The low geometric resolution HSRI can be modelled as a geometrically downsampled version of $\mathbf{Z}$:

$$\mathbf{H} \approx \mathbf{ZS} = \mathbf{EAS} = \mathbf{E\tilde{A}} , \qquad (4)$$

where $\mathbf{S} \in \mathbb{R}^{N_m \times N_h}$ is the downsampling operator that describes the spatial response of the hyperspectral sensor, and $\tilde{\mathbf{A}} \equiv \mathbf{AS}$ are the abundances at the lower resolution. $\mathbf{S}$ can be approximated by taking the (weighted) average values of the multispectral image corresponding to one HSRI pixel.

Likewise, the low spectral resolution HGRI can be modelled as a spectral downsampling of $\mathbf{Z}$

$$\mathbf{M} \approx \mathbf{RZ} = \mathbf{REA} = \tilde{\mathbf{E}}\mathbf{A} , \qquad (5)$$

where $\mathbf{R} \in \mathbb{R}^{b \times B}$ is the spectral response function of the multispectral sensor and each row contains the responses of each multispectral band. $\tilde{\mathbf{E}} \equiv \mathbf{RE}$ are the spectrally transformed (downsampled) endmembers. For this work, we assume that $\mathbf{S}$ and $\mathbf{R}$ are perfectly known.

4. PROPOSED SOLUTION

The fusion is solved if $\mathbf{Z}$ has been found, which is the same as estimating $\mathbf{E}$ and $\mathbf{A}$. Actually, by solving for $\mathbf{E}$ and $\mathbf{A}$ separately we are solving first the spectral unmixing and then using the result to generate the fused image $\mathbf{Z}$. By taking into account the constraints (3), we formulate the following constrained least-squares problem:

$$\arg\min_{\mathbf{E},\mathbf{A}} \ \|\mathbf{H} - \mathbf{EAS}\|_F^2 + \|\mathbf{M} - \mathbf{REA}\|_F^2 \qquad (6a)$$

$$\text{subject to } 0 \leq e_{ij} \leq 1, \quad \forall i,j \qquad (6b)$$

$$a_{ij} \geq 0, \quad \forall i,j \qquad (6c)$$

$$\mathbf{1}^\top \mathbf{A} = \mathbf{1}^\top \qquad (6d)$$

with $\| \cdot \|_F$ denoting the Frobenius norm. The non-negativity and sum-to-one constraints (6c, 6d) geometrically mean that the abundances are forced to lie on the surface of a simplex whose vertices are the endmembers. This property at the same time acts as a prior sparsity term, which promotes that in each pixel only a limited number of endmembers should be present.

Solving (6a) for $\mathbf{E}$ and $\mathbf{A}$ is rather tricky and suffers from the following issues. The first term is ill-posed with respect to $\mathbf{A}$, because the downsampling operator $\mathbf{S}$ acts on the spatial domain and any update coming to $\mathbf{A}$ from this term will suffer from the low geometric resolution of the HSRI. In much the same way, the second term is ill-posed with respect to $\mathbf{E}$, because the HGRI has broad spectral responses $\mathbf{R}$ and only coarse information about the material spectra can be acquired. Consequently, we propose to split (6a) into two subproblems and solve them by alternating between the (geometrically) low-resolution term for $\mathbf{H}$ and the high-resolution term for $\mathbf{M}$. This is somewhat similar to Yokoya et al. (2012), although they omit constraints other than non-negativity (but do a full matrix factorization in each step rather than updating only endmembers *or* the abundances).

Solving the first term of (6a) subject to the constraints on $\mathbf{E}$ will be referred to as the low-resolution step:

$$\arg\min_{\mathbf{E}} \ \|\mathbf{H} - \mathbf{E\tilde{A}}\|_F^2$$
$$\text{subject to } 0 \leq e_{ij} \leq 1, \quad \forall i,j . \qquad (7)$$

Here, the endmembers $\mathbf{E}$ are updated based on the low-resolution abundances $\tilde{\mathbf{A}}$, which are kept fixed. The minimisation of the second term of (6a) under the constraints on $\mathbf{A}$ is the high-resolution step:

$$\arg\min_{\mathbf{A}} \ \|\mathbf{M} - \tilde{\mathbf{E}}\mathbf{A}\|_F^2$$
$$\text{subject to } a_{ij} > 0, \quad \forall i,j \qquad (8)$$
$$\mathbf{1}^\top \mathbf{A} = \mathbf{1}^\top$$

The high-resolution abundances $\mathbf{A}$ are updated and the spectrally downsampled endmembers $\tilde{\mathbf{E}}$ acquired from the low-resolution step are kept fixed.

Algorithm 1 Solution of minimization problem (6a).

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 (9a, 9b)
    // high-resolution step:
    $\tilde{\mathbf{E}} \leftarrow \mathbf{R}\mathbf{E}^{(k)}$; Estimate $\mathbf{A}^{(k)}$ with (10a, 10b)
  end while
  return $\mathbf{Z} = \mathbf{E}^{(k)}\mathbf{A}^{(k)}$

4.1 Optimisation scheme

Problems (7, 8) are both constrained least-squares and we alternate between the two. We propose to use a projected gradient method to solve both, inspired by the PALM (Proximal Alternating Linearised Minimization) algorithm (Bolte et al., 2014). For (7) the following two steps are iterated until convergence for a number of iterations $q = 1, 2, ...$:

$$\mathbf{U}^q = \mathbf{E}^{q-1} - \frac{1}{c}(\mathbf{E}^{q-1}\tilde{\mathbf{A}} - \mathbf{H})\tilde{\mathbf{A}}^\top \quad (9a)$$

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

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 (7). The proximal operator simply truncates the entries of $\mathbf{U}$ to 0 from below and to 1 from above. Parameter γ_1 is explained below.

Similarly, (8) is minimised by iterating the following steps until convergence:

$$\mathbf{V}^q = \mathbf{A}^{q-1} - \frac{1}{d}\tilde{\mathbf{E}}^\top(\tilde{\mathbf{E}}\mathbf{A}^{q-1} - \mathbf{M}) \quad (10a)$$

$$\mathbf{A}^q = \text{prox}_{\mathbf{A}}(\mathbf{V}^q) \quad (10b)$$

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 (8). For this simplex projection, a recent, computationally efficient algorithm is available (Condat, 2014). We set $\gamma_1 = \gamma_2 = 1.01$ for the experiments. These parameters only affect the speed of convergence. An overview over the whole optimisation procedure is given in Algorithm 1.

The problem described in (6a) is highly non-convex and the local optimisation depends a lot on the initial values. To acquire good initial values for the endmembers we use SISAL (Bioucas-Dias, 2009). SISAL is based on the principle that the simplex defined by the endmembers must have minimum volume. It uses a sequence of augmented Lagrangian optimizations in a robust way to return the endmembers. Once the inital endmembers $\mathbf{E}^{(0)}$ are defined, we then use SUnSAL (Bioucas-Dias and Nascimento, 2008) to get initial abundances. SUnSAL solves a least-squares problem for $\tilde{\mathbf{A}}^{(0)}$ via ADMM and can incorporate many constraints, but we only use the constraints (6c, 6d). Finally, we initialize the high geometric resolution abundances $\mathbf{A}^{(0)}$, by upsampling them with the pseudo-inverse of the downsampling operator, $\mathbf{A}^{(0)} = \tilde{\mathbf{A}}^{(0)}(\mathbf{S}^\top\mathbf{S})^{-1}\mathbf{S}^\top$. We observe that strict upsampling produces to blocking artefacts, so we apply a low-pass filter to $\mathbf{A}^{(0)}$.

5. EXPERIMENTS

5.1 Datasets

Acquiring HSRI and HGRI simultaneously is not yet a standard procedure, and even if such data were available extensive ground truth would be needed to be able to evaluate the fusion method. Consequently, in this study we make use of semi-synthetic data. That means that as input we have only one real hyperspectral image (which we call the "ground truth"), from which we generate a HSRI and a HGRI by geometric, respectively spectral downsampling. These HSRI and HGRI are synthetic, but since they are based on a true HSRI and realistic spectral response functions, we call them *semi*-synthetic data.

We use in total four publically available hyperspectral images to validate our methodology. First, two well-known images, Cuprite (in Figure 1) and Indian Pines[1], were selected as ground truth. These were acquired with AVIRIS, a sensor of NASA that delivers images with 224 contiguous spectral bands in the 0.4-$2.5\mu m$ range (Varshney and Arora, 2004). After removing the water absorption bands, the images had 195 and 199 spectral bands, respectively. The Cuprite image has dimensions 512×608 pixels and Indian Pines 145×145 pixels. The respective GSDs are 18 m and 10 m.

Additionally, we use another standard hyperspectral image, *Pavia University*[2], which was acquired with ROSIS, a sensor of DLR [3] that has 115 spectral bands spanning the 0.4-$0.9\mu m$ range. The image has dimentions 610×340 pixels and 103 spectral bands are available, with an approximate GSD of 1 m.

Finally, we choose to use an image from APEX (Schaepman et al., 2015), developed by a Swiss-Belgian consortium on behalf of ESA. APEX covers the spectral range 0.4-$2.5\mu m$. The image has a GSD of about 2 m and was acquired in 2011 over Baden, Switzerland (true color composite in Figure 4), as an Open Science Dataset[4]. The cropped image we used has dimensions 400×400 pixels and 211 spectral bands, after removing water absorption bands.

All images described above serve as ground truth for our experiments. To simulate the HSRI $\mathbf{H}$ from each image we downsample them with a factor of 8, which is a realistic difference between observed HSRIs and HGRIs. The downsampling is done by simple averaging over 8×8 pixel blocks. The HGRI $\mathbf{M}$ can be simulated using the known spectral responses $\mathbf{R}$ of existing multispectral sensors. We choose to use a different spectral response $\mathbf{R}$ for each hyperspectral sensor based on the geometric resolution of the HSRIs. Thus, we take the spectral response of Landsat 8 to combine with the scenes acquired with AVIRIS, because of the moderate geometric resolution. Secondly, the IKONOS spectral response is used for the ROSIS sensor scene, since the HSRI only covers the VNIR spectrum. Finally we we choose the spectral response of the Leica ADS80 for the scene acquired with APEX. The last case will be treated as more challenging, since the two spectra do not fully overlap. We take care that the area under the spectral response for each band is normalized to 1, so that the simulated HGRI has the same value range as the HSRI.

---

[1] Available from: `http://aviris.jpl.nasa.gov/data/free_data.html`, `https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html`

[2] Available from: `http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes`

[3] More information at: `http://messtec.dlr.de/en/technology/dlr-remote-sensing-technology-institute/hyperspectral-systems-airborne-rosis-hyspex/`

[4] Available from: `http://www.apex-esa.org/content/free-data-cubes`

Method	APEX			Pavia University			Cuprite			Indian Pines		
	RMSE	ERGAS	SAM	RMSE	ERGAS	SAM	RMSE	ERGAS	SAM	RMSE	ERGAS	SAM
Wycoff	8.47	3.03	7.57	2.32	0.83	2.66	1.64	0.18	0.69	2.98	0.70	2.39
Akhtar	14.40	5.38	12.81	4.48	1.47	4.76	7.24	0.74	2.99	3.39	0.77	2.66
HySure	10.58	3.97	10.55	3.00	1.02	3.11	2.02	0.21	0.84	3.34	0.79	2.68
Ours	**7.88**	**2.86**	**6.72**	**2.28**	**0.80**	**2.37**	**1.60**	**0.17**	**0.67**	**2.88**	**0.69**	**2.29**

Table 1. Quantitative results for all tested images.

5.2 Implementation details and baseline methods

We run our method with the maximum number of endmembers set to $p = 30$. Although lower numbers would be enough for most of the images, we choose the number conservatively to adapt to the big variability within the APEX scene. This might cause the algorithm to overfit in some cases, but does not seem to affect the quantitative results. The inner loops of the two optimisation steps (9a), (10a) run quite fast and typically converge in ≈ 10 iterations in the early stages, dropping to ≈ 2 iterations as the alternation proceeds. Inner loops stop when the update falls below 1%. We iterate the outer loop over (7, 8) until the overall cost (6a) changes less than 0.01%, or for at most 2000 iterations. For the APEX dataset the proposed method took ≈ 10 min to run in a Matlab implementation on a single Intel Xeon E5 3.2 GHz CPU. Note, for a given image the processing speed depends heavily on the chosen number p of endmembers.

We use three state-of-the-art fusion methods (Wycoff et al., 2013, Akhtar et al., 2014, Simoes et al., 2015) as baselines to compare against. Below, we will call the first two (nameless) methods *Wycoff* and *Akhtar* and the third one *HySure*, the name coined by the authors. The first two methods come from the computer vision literature, where they reported the best results (especially for close-range images), while the third one comes from recent remote sensing literature. We use the authors' original implementations, which are available for all three baselines. All methods were run with the exact same spectral responses **R**. For HySure, we use only the fusion part and deactivate the additional estimation of the point spread function. We treat the latter as perfectly known, so re-estimating it would give the method a disadvantage. Since we used $p = 30$ endmembers for our method, we also used a subspace of 30 for HySure. For the method of Akhtar we adapted the number of atoms that is learnt in the dictionary to $k = 200$, so that the dictionary can cover more spectral responses. Finally, for the method of Wycoff we could not increase the number of scene materials from $N = 10$, the initial number proposed. Even with $N = 10$ the method needed more than 100 GB of memory to perform the fusion.

5.3 Error metrics

As primary error metric, the root mean square error of the estimated high-resolution hyperspectral image $\hat{\mathbf{Z}}$ compared to the ground truth $\mathbf{Z}$ is used, scaled to an 8-bit intensity range $[0 \ldots 255]$.

$$\text{RMSE} = \sqrt{\frac{1}{BN_m} \sum \sum (\hat{z}_{ij} - z_{ij})^2} = \sqrt{\frac{\|\hat{\mathbf{Z}} - \mathbf{Z}\|_F^2}{BN_m}} \quad (11)$$

To make the evaluation independent of intensity units, we additionally use the Erreur Relative Globale Adimensionnelle de Synthèse (Wald, 2000). This error measure is also supposed to be independent of the GSD difference between the HSRI and HGRI.

$$\text{ERGAS} = 100S \sqrt{\frac{1}{B} \sum \frac{\text{MSE}(\hat{\mathbf{z}}_i, \mathbf{z}_i)}{\mu_{\hat{\mathbf{z}}_i}^2}}, \quad (12)$$

where S is the ratio of GSD difference of the HGRI and HSRI ($S = \text{GSD}_h / \text{GSD}_m$), $\text{MSE}(\hat{\mathbf{z}}_i, \mathbf{z}_i)$ is the mean squared error of every estimated spectral band $\hat{\mathbf{z}}_i$ and $\mu_{\hat{\mathbf{z}}_i}$ is the mean value of each spectral band. Finally, we used the Spectral Angle Mapper (SAM, Yuhas et al., 1992), which is the average, among all pixels, of the angle in $\mathbb{R}^B$ between the estimated pixel $\hat{\mathbf{z}}_j$ and the ground truth pixel $\mathbf{z}_j$.

$$\text{SAM} = \frac{1}{N_m} \sum \arccos \frac{\hat{\mathbf{z}}_j^\top \mathbf{z}_j}{\|\hat{\mathbf{z}}_j\|_2 \|\mathbf{z}_j\|_2} \quad (13)$$

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

6. EXPERIMENTAL RESULTS AND DISCUSSION

Table 1 shows the error metrics (in bold best results) that each method achieved for all four images. The proposed fusion approach with joint unmixing performs better than all other methods, across all three error measures. The improvement can be seen especially in terms of RMSE and SAM. The effect of resolution difference (i.e. the ratio S) in ERGAS reduces the differences of the methods. Among the other methods, the method of Wycoff performs the best. The method of Akhtar, proposed mainly for close-range computer vision images, performs worse for remote sensing images. Mostly the best performance is achieved for Cuprite, while for APEX all methods perform worst.

Complementary to the tabulated results, we visualize the per-pixel residuals for all four investigated images. In Figure 2 the pixels are sorted in ascending order according to their per-pixel RMSE, computed over all spectral bands. From this graph it can be seen that our method in all cases has the largest amount of pixels with small reconstruction errors. E.g. in images Pavia University and Cuprite most of the pixels are reconstructed with an RMSE lower than 2 gray values in 8bit range. We also visualize the RMSE per spectral band in Figure 3. Based on this figure, our method achieves very similar results with the method of Wycoff, except in the case of APEX, where the proposed method outperforms the baseline, especially in the wavelengths above 1000 nm. The method of Akhtar shows big and rapid RMSE variations, depending on the wavelength. It also shows a different shape of the spectral response curves, compared to the other three methods that have quite similar shapes. In APEX we note rapidly growing RMSE for wavelengths above 900 nm. This is due to the fact that ADS80 has no spectral bands in this range.

Moreover, we choose to visualize the error in two bands of the APEX dataset for all methods. We choose APEX because it is the most challenging scene, since contains many different materials in a mixed urban scene and also has the largest reconstruction errors. Moreover, this image gives bigger differences between the methods. In Figure 4, we visualize the differences (in 8 bit range) between the ground truth and reconstructed images for wavelengths 552 nm and 1506 nm, corresponding to band with low reconstruction error (spectral overlap with the HGRI), respectively high reconstruction error (see Figure 3a).
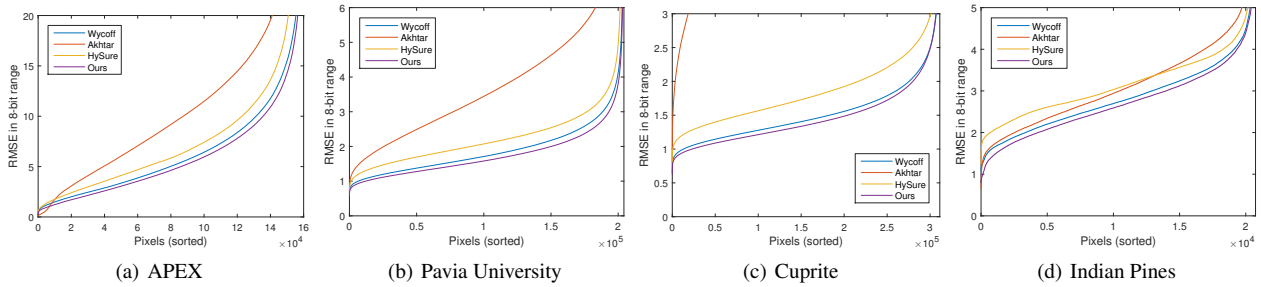
(a) APEX   (b) Pavia University   (c) Cuprite   (d) Indian Pines

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

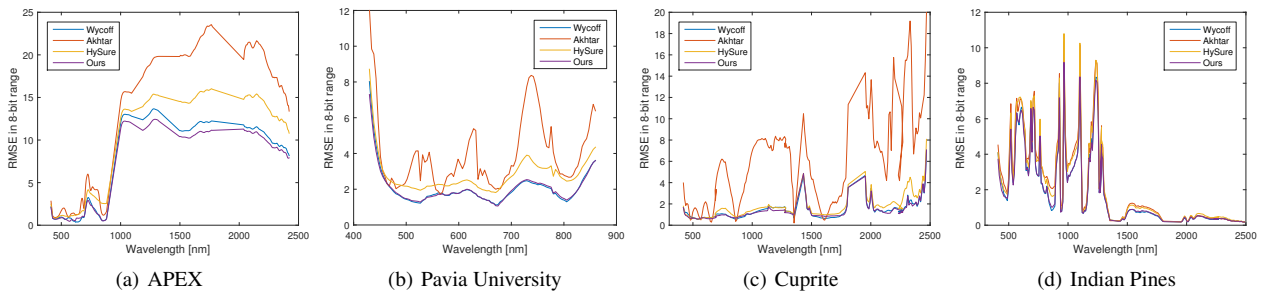(a) APEX   (b) Pavia University   (c) Cuprite   (d) Indian Pines

Figure 3. Per-band RMSE for every spectral image.

The evaluation shows that we were able to generate a fused image that is 8×8 times improved in the geometric resolution compared to the input HSRI. At the same time, by using the joint spectral unmixing we are able to acquire a physically plausible explanation of the scene's endmembers and fractional abundances. Figure 5 shows the abundances for the APEX scene corresponding to the endmembers (a) for the surface of a tennis court and (b) for stadium tartan. On the top, on the left side is the abundance map for the fused image as recovered by our method, while on the right is the abundance map for the HSRI. Clearly the HGRI provides information about the spatial distribution of the endmembers within a hyperspectral image pixel. This indicates that we can improve unmixing results of the HSRI. However, we are not able to quantitatively judge the unmixing quality, as no ground truth exists. Note that in the true color image of Figure 4 the tennis court and the stadium tartan have very similar color. However, the abundance maps in Figure 5 for the tennis court and the stadium tartan endmembers show different values for these two objects, which should be so, since these two objects consist of different materials. In Figure 5 (bottom right) are the average spectral responses of 30 representative pixels of each material, taken from the ground truth image. Although these two materials reflect similarly in the visible range, they have considerable differences in the infrared range.

The method we proposed has only one essential user parameter, namely the number p of endmembers. If fusion is the ultimate goal, then preliminary tests have proven that it is better to choose p larger than the actual number of endmembers in the scene. In this case, probably some of the estimated endmembers correspond to the same material, splitting the spectral response into two parts (i.e., endmember is no longer synonymous with surface material). Nevertheless this will not affect the fused image. More endmembers may also help handling shadows and specular reflections.

We believe that the most important ingredient for the success of our method is the simplex constraint (6c, 6d), that greatly stabilizes the model. Although these two constraints by themselves

are standard for spectral unmixing, we are not aware of any other method that uses both of them together for image fusion.

In this paper, we used semi-synthetic, perfectly co-registered data. Under these circumstances, most methods are able to reconstruct the images with RMSE $\approx 1\%$ of the intensity range, except for APEX. In case that the HSRI and HGRI were acquired by different platforms, co-registration errors between the two images would affect the fusion and the images would have to be corrected. Furthermore, different sensors will have different offsets and gains and the spectral response $\mathbf{R}$ must be modified so that the value range for all images is the same. The situation gets even more complicated if the images are not taken simultaneously, when multitemporal effects will influence the fusion, especially illumination differences.

As a closing remark, some images studied in this work have important deficiencies. Images from the AVIRIS sensor have a lot of noise, especially the Indian Pines image. In Figure 3d the peaks in graphs indicate that in those bands there is a lot of noise, which cannot be explained by any method. Recent sensors like APEX seem not to have unusually noisy bands and it is important that new challenging hyperspectral test datasets are provided to the community to help improve algorithms and applications.

7. CONCLUSIONS

In this work, we have descussed synergies between hyperspectral fusion and unmixing, some investigated for the first time, to the best of our knowledge. First, we perform an integrated processing of both HSRI and HGRI to jointly solve the tasks of image fusion and spectral unmixing. Thereby, the spectral unmixing supports image fusion and vice versa. Second, in the joint estimation, we enforce several useful physical constraints simultaneously, which lead to a more plausible estimation of endmembers and abundances, and at the same time stabilize fusion. Third, we adopt efficient recent optimization machinery which makes it possible to tackle the resulting estimation problem. The validity of the proposed method has been shown experimentally in tests
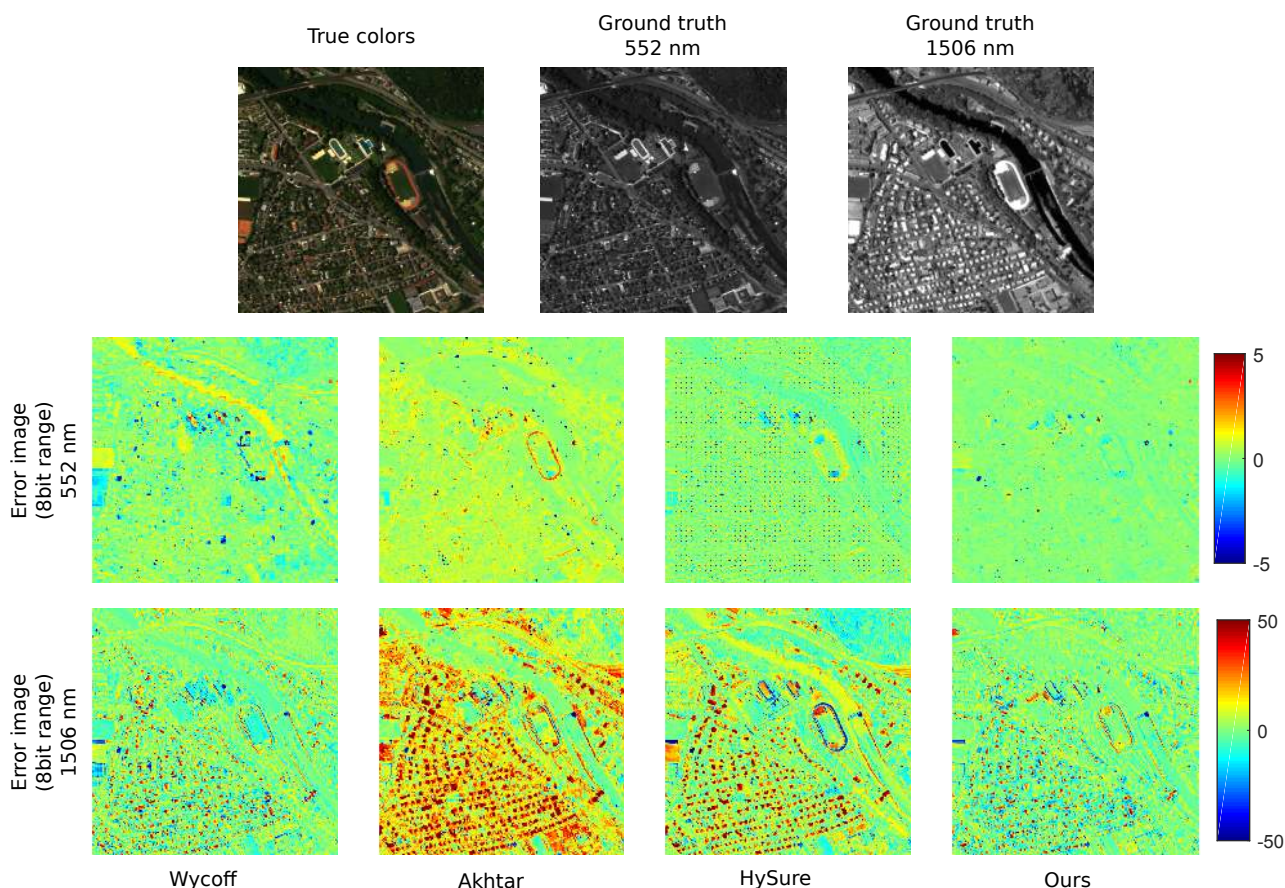
Figure 4. *Top row:* True color image for APEX and ground truth spectral images for wavelengths 552 and 1506 nm. *Second and third rows:* The reconstruction error for each method at the mentioned wavelengths. Note the difference in the color scale between the two rows.
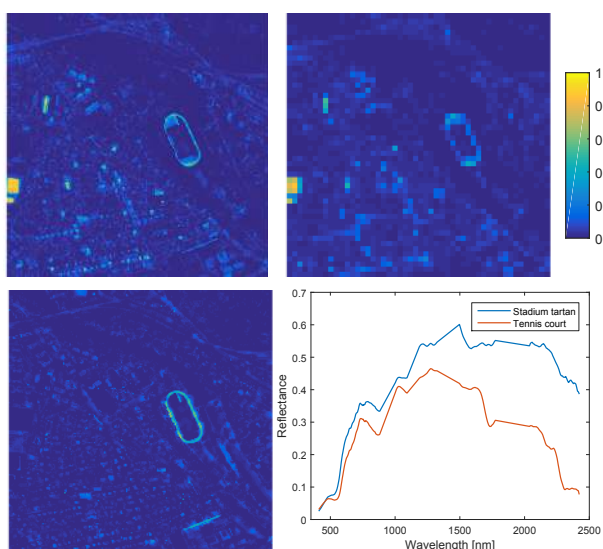

Figure 5. *Top row:* Abundance maps corresponding to the end-member of a tennis court surface: for the fused image (left), for the HSRI (right). *Bottom row:* Abundance map for the fused image corresponding to the stadium tartan endmember (left), the ground truth spectra of the respective materials (right).

with four semi-synthetic datasets, including a comparison to three state-of-the-art fusion methods. In the tests our method improved the RMSE of tyhe fusion result by 4% to 53% relative to the best and worst baseline, respectively.

Future work will include several aspects: Processing of real HSRI and HGRI data, and comparison to external ground truth; Scaling to larger scenes, potentially with a locally varying set of endmembers so as not to unnecessarily inflate the spectral basis; Better modelling of multitemporal differences and shadows, which account for an important part of the error in the current processing; Extension to multi-layer unmixing that occurs in vegetation applications when tree canopies interact with each other and with the ground.

ACKNOWLEDGEMENTS

REFERENCES

Akhtar, N., Shafait, F. and Mian, A., 2014. Sparse spatio-spectral representation for hyperspectral image super-resolution. In: Computer Vision–ECCV 2014, Springer, Heidelberg, pp. 63–78.

Ben-Dor, E., Schläpfer, D., Plaza, A. J. and Malthus, T., 2013. Hyperspectral remote sensing. In: Airborne Measurements for Environmental Research, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, pp. 413–456.

Bioucas-Dias, J. M., 2009. A variable splitting augmented lagrangian approach to linear spectral unmixing. In: First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS'09), IEEE, pp. 1–4.

Bioucas-Dias, J. M. and Nascimento, J. M., 2008. Hyperspectral subspace identification. IEEE Trans. on Geoscience and Remote Sensing 46(8), pp. 2435–2445.

Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N. M. and Chanussot, J., 2013. Hyperspectral remote sensing data analysis and future challenges. IEEE, Geoscience and Remote Sensing Magazine 1(2), pp. 6–36.

Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P. and Chanussot, J., 2012. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5(2), pp. 354–379.

Bolte, J., Sabach, S. and Teboulle, M., 2014. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Mathematical Programming 146(1-2), pp. 459–494.

Chan, T.-H., Ma, W.-K., Ambikapathi, A. and Chi, C.-Y., 2011. A simplex volume maximization framework for hyperspectral endmember extraction. IEEE Trans. on Geoscience and Remote Sensing 49(11), pp. 4177–4193.

Condat, L., 2014. Fast projection onto the simplex and the L1 ball. [Online] http://hal.univ-grenoble-alpes.fr/hal-01056171/, Accessed: 16.03.2015.

Gu, Y., Zhang, Y. and Zhang, J., 2008. Integration of spatial–spectral information for resolution enhancement in hyperspectral images. IEEE Trans. on Geoscience and Remote Sensing 46(5), pp. 1347–1358.

Hardie, R. C., Eismann, M. T. and Wilson, G. L., 2004. Map estimation for hyperspectral image resolution enhancement using an auxiliary sensor. IEEE Trans. on Image Processing 13(9), pp. 1174–1184.

Heinz, D. C. and Chang, C.-I., 2001. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. IEEE Trans. on Geoscience and Remote Sensing 39(3), pp. 529–545.

Heylen, R., Parente, M. and Gader, P., 2014. A review of nonlinear hyperspectral unmixing methods. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(6), pp. 1844–1868.

Huang, B., Song, H., Cui, H., Peng, J. and Xu, Z., 2014. Spatial and spectral image fusion using sparse matrix factorization. IEEE Trans. on Geoscience and Remote Sensing 52(3), pp. 1693–1704.

Iordache, M.-D., Bioucas-Dias, J. M. and Plaza, A., 2012. Total variation spatial regularization for sparse hyperspectral unmixing. IEEE Trans. on Geoscience and Remote Sensing 50(11), pp. 4484–4502.

Kasetkasem, T., Arora, M. K. and Varshney, P. K., 2005. Super-resolution land cover mapping using a markov random field based approach. Remote Sensing of Environment 96(3), pp. 302–314.

Kawakami, R., Wright, J., Tai, Y.-W., Matsushita, Y., Ben-Ezra, M. and Ikeuchi, K., 2011. High-resolution hyperspectral imaging via matrix factorization. In: IEEE Conf. Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 2329–2336.

Keshava, N. and Mustard, J. F., 2002. Spectral unmixing. IEEE Signal Processing Magazine 19(1), pp. 44–57.

Miao, L. and Qi, H., 2007. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. IEEE Trans. on Geoscience and Remote Sensing 45(3), pp. 765–777.

Nascimento, J. M. and Bioucas Dias, J. M., 2005. Vertex component analysis: A fast algorithm to unmix hyperspectral data. IEEE Trans. on Geoscience and Remote Sensing 43(4), pp. 898–910.

Nascimento, J. M. and Bioucas-Dias, J. M., 2012. Hyperspectral unmixing based on mixtures of dirichlet components. IEEE Trans. on Geoscience and Remote Sensing 50(3), pp. 863–878.

Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J. C. and Trianni, G., 2009. Recent advances in techniques for hyperspectral image processing. Remote Sensing of Environment 113, Supplement 1, pp. S110–S122.

Schaepman, M. E., Jehle, M., Hueni, A., D'Odorico, P., Damm, A., Weyermann, J., Schneider, F. D., Laurent, V., Popp, C., Seidel, F. C., Lenhard, K., Gege, P., Kuechler, C., Brazile, J., Kohler, P., Vos, L. D., Meuleman, K., Meynart, R., Schlaepfer, D., Kneubuehler, M. and Itten, K. I., 2015. Advanced radiometry measurements and earth science applications with the airborne prism experiment (APEX). Remote Sensing of Environment 158, pp. 207–219.

Simoes, M., Bioucas-Dias, J., Almeida, L. and Chanussot, J., 2015. A convex formulation for hyperspectral image superresolution via subspace-based regularization. IEEE Trans. on Geoscience and Remote Sensing 53(6), pp. 3373–3388.

Varshney, P. K. and Arora, M. K., 2004. Advanced image processing techniques for remotely sensed hyperspectral data. Springer, Berlin.

Wald, L., 2000. Quality of high resolution synthesised images: Is there a simple criterion? In: Third conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images", SEE/URISCA, pp. 99–103.

Wei, Q., Dobigeon, N. and Tourneret, J.-Y., 2014. Bayesian fusion of hyperspectral and multispectral images. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp. 3176–3180.

Wycoff, E., Chan, T.-H., Jia, K., Ma, W.-K. and Ma, Y., 2013. A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp. 1409–1413.

Yokoya, N., Yairi, T. and Iwasaki, A., 2012. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. IEEE Trans. on Geoscience and Remote Sensing 50(2), pp. 528–537.

Yuhas, R. H., Goetz, A. F. and Boardman, J. W., 1992. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In: Summaries of the Third Annual JPL Airborne Geoscience Workshop, Vol. 1, pp. 147–149.

Zare, A. and Gader, P., 2007. Sparsity promoting iterated constrained endmember detection in hyperspectral imagery. IEEE Geoscience and Remote Sensing Letters 4(3), pp. 446–450.