Sparse Spatio-spectral Representation for Hyperspectral Image Super-Resolution

Naveed Akhtar, Faisal Shafait and Ajmal Mian

School of Computer Science and Software Engineering.
The University of Western Australia.
35 Stirling Highway, 6009 Crawley WA.
naveed.akhtar@research.uwa.edu.au, {faisal.shafait, ajmal.mian} @uwa.edu.au

Abstract. Existing hyperspectral imaging systems produce low spatial resolution images due to hardware constraints. We propose a sparse representation based approach for hyperspectral image super-resolution. The proposed approach first extracts distinct reflectance spectra of the scene from the available hyperspectral image. Then, the signal sparsity, non-negativity and the spatial structure in the scene are exploited to explain a high-spatial but low-spectral resolution image of the same scene in terms of the extracted spectra. This is done by learning a sparse code with an algorithm G-SOMP+. Finally, the learned sparse code is used with the extracted scene spectra to estimate the super-resolution hyperspectral image. Comparison of the proposed approach with the state-of-the-art methods on both ground-based and remotely-sensed public hyperspectral image databases shows that the presented method achieves the lowest error rate on all test images in the three datasets.

Keywords: Hyperspectral, super-resolution, spatio-spectral, sparse representation.

1 Introduction

Hyperspectral imaging acquires a faithful representation of the scene radiance by integrating it against several basis functions that are well localized in the spectral domain. The spectral characteristics of the resulting representation have proven critical in numerous applications, ranging from remote sensing [1], [2] to medical imaging [3]. They have also been reported to improve the performance in computer vision tasks, such as, tracking [4], segmentation [5], recognition [6] and document analysis [7]. However, contemporary hyperspectral imaging lacks severely in terms of spatial resolution [3], [8]. The problem stems from the fact that each spectral image acquired by a hyperspectral system corresponds to a *very narrow* spectral window. Thus, the system must use long exposures to collect enough photons to maintain a good signal-to-noise ratio of the spectral images. This results in low spatial resolution of the hyperspectral images.

Normally, spatial resolution can be improved with high resolution sensors. However, this solution is not too effective for hyperspectral imaging, as it further

reduces the density of the photons reaching the sensor. Keeping in view the hardware limitations, it is highly desirable to develop software based techniques to enhance the spatial resolution of hyperspectral images. In comparison to the hyperspectral systems, the low spectral resolution imaging systems (e.g. RGB cameras) perform a gross quantization of the scene radiance - loosing most of the spectral information. However, these systems are able to preserve much finer spatial information of the scenes. Intuitively, images acquired by these systems can help in improving the spatial resolution of the hyperspectral images.

This work develops a sparse representation [9] based approach for hyperspectral image super-resolution, using a high-spatial but low-spectral resolution image (henceforth, only called the *high spatial resolution image*) of the same scene. The proposed approach uses the hyperspectral image to extract the reflectance spectra related to the scene. This is done by solving a constrained sparse representation problem using the hyperspectral image as the input. The basis formed by these spectra is transformed according to the spectral quantization of the high spatial resolution image. Then, the said image and the transformed basis are fed to a simultaneous sparse approximation algorithm G-SOMP+. Our algorithm is a generalization of Simultaneous Orthogonal Matching Pursuit (SOMP) [10] that additionally imposes a non-negativity constraint over its solution space. Taking advantage of the spatial structure in the scene, G-SOMP+ efficiently learns a sparse code. This sparse code is used with the reflectance spectra of the scene to estimate the super-resolution hyperspectral image. We test our approach using the hyperspectral images of objects, real-world indoor and outdoor scenes and remotely sensed hyperspectral image. Results of the experiments show that the proposed approach consistently performs better than the existing methods on all the data sets.

This paper is organized as follows. Section 2 reviews the previous literature related to the proposed approach. We formalize our problem in Section 3. The proposed solution is described in Section 4 of the paper. In Section 5, we give the results of the experiments that have been performed to evaluate the approach. We dedicate Section 6 for the discussion on the results and the parameter settings. The paper concludes with a brief summary in Section 7.

2 Related work

Hardware limitations have lead to a notable amount of research in software based techniques for high spatial resolution hyperspectral imaging. The software based approaches that use image fusion [11] as a tool, are particularly relevant to our work. Most of these approaches have originated in the remote sensing literature because of the early introduction of hyperspectral imaging in the airborne/spaceborne observatory systems. In order to enhance the spatial resolution of the hyperspectral images, these approaches usually fuse a hyperspectral image with a high spatial resolution pan-chromatic image. This process is known as pan-sharpening [12]. A popular technique ([13], [14], [15], [16]) uses a linear transformation of the color coordinates to improve the spatial resolution

of hyperspectral images. Exploiting the fact that human vision is more sensitive to luminance, this technique fuses the luminance component of a high resolution image with the hyperspectral image. Generally, this improves the spatial resolution of the hyperspectral image, however the resulting image is sometimes spectrally distorted [17].

In spatio-spectral image fusion, one class of methods exploits spatial unmixing ([18], [19]) for improving the spatial resolution of the hyperspectral images. These methods only perform well for the cases when the spectral resolutions of the two images are not very different. Furthermore, their performance is compromised in highly mixed scenarios [8]. Zurita-Milla et al. [20] employed a sliding window strategy to mitigate this issue. Image filtering is also used for interpolating the spectral images to improve the spatial resolution [21]. In this case, the implicit assumption of smooth spatial patterns in the scenes often produces overly smooth images.

More recently, matrix factorization has played an important role in enhancing the spatial resolution of the ground based and the remote sensing hyperspectral imaging systems ([3], [8], [22], [23]). Kawakami et al. [3] have proposed to fuse a high spatial resolution RGB image with a hyperspectral image by decomposing each of the two images into two factors and constructing the desired image from the complementary factors of the two decompositions. A very similar technique has been used by Huang et al. [8] for remote sensing data. The main difference between [3] and [8] is that the latter uses a spatially down-sampled version of the high spatial resolution image in the matrix factorization process. Wycoff et al. [22] have proposed an algorithm based on Alternating Direction Method of Multipliers (ADMM) [24] for the factorization of the matrices and later using it to fuse the hyperspectral image with an RGB image. Yokoya et al. [23] have proposed a coupled matrix factorization approach to fuse multi-spectral and hyperspectral remote sensing images to improve the spatial resolution of the hyperspectral images.

The matrix factorization based methods are closely related to our approach. However, our approach has major differences with each one of them. Contrary to these methods, we exploit the spatial structure in the high spatial resolution image for the improved performance. The proposed approach also takes special care of the physical significance of the signals and the processes related to the problem. This makes our formalization of the problem and its solution unique. We make use of the non-negativity of the signals, whereas [3] and [8] do not consider this notion at all. In [22] and [23] the authors do consider the non-negativity of the signals, however their approaches require *a priori* knowledge of the spatial transform between the input hyperspectral image and the input high spatial resolution image. This requirement compromises the practicality of these approaches. Our approach does not does not impose any such requirement.

3 Problem formulation

We seek estimation of a super-resolution hyperspectral image $\mathbf{S} \in \mathbb{R}^{M \times N \times L}$, where M and N denote the spatial dimensions and L represents the spectral

dimension, from an acquired hyperspectral image $\mathbf{Y}_h \in \mathbb{R}^{m\times n\times L}$ and a corresponding high spatial (but low spectral) resolution image of the same scene $\mathbf{Y} \in \mathbb{R}^{M\times N\times l}$. For our problem, $m \ll M, n \ll N$ and $l \ll L$, which makes the problem severely ill-posed. We consider both of the available images to be linear mappings of the target image:

$$\mathbf{Y} = \Psi(\mathbf{S}), \quad \mathbf{Y}_h = \Psi_h(\mathbf{S}) \tag{1}$$

where, $\Psi : \mathbb{R}^{M\times N\times L} \rightarrow \mathbb{R}^{M\times N\times l}$ and $\Psi_h : \mathbb{R}^{M\times N\times L} \rightarrow \mathbb{R}^{m\times n\times L}$.

A typical scene of the ground based imagery as well as the space-borne/airborne imagery contains only a small number of distinct materials [3], [25]. If the scene contains q materials, the linear mixing model (LMM) [26] can be used to approximate a pixel $\mathbf{y}_h \in \mathbb{R}^L$ of $\mathbf{Y}_h$ as

$$\mathbf{y}_h \approx \sum_{\omega=1}^{c} \boldsymbol{\varphi}_\omega \alpha_\omega \quad , \quad c \leq q \tag{2}$$

where, $\boldsymbol{\varphi}_\omega \in \mathbb{R}^L$ denotes the reflectance of the ω-th distinct material in the scene and α_ω is the *fractional abundance* (i.e. proportion) of that material in the area corresponding to the pixel. We rewrite (2) in the following matrix form:

$$\mathbf{y}_h \approx \boldsymbol{\Phi}\boldsymbol{\alpha} \tag{3}$$

In (3), the columns of $\boldsymbol{\Phi} \in \mathbb{R}^{L\times c}$ represent the reflectance vectors of the underlying materials and $\boldsymbol{\alpha} \in \mathbb{R}^c$ is the coefficient vector. Notice that, when the scene represented by a pixel $\mathbf{y}_h$ also includes the area corresponding to a pixel $\mathbf{y} \in \mathbb{R}^l$ of $\mathbf{Y}$, we can approximate $\mathbf{y}$ as

$$\mathbf{y} \approx (\mathbf{T}\boldsymbol{\Phi})\boldsymbol{\beta} \tag{4}$$

where, $\mathbf{T} \in \mathbb{R}^{l\times L}$ is a transformation matrix and $\boldsymbol{\beta} \in \mathbb{R}^c$ is the coefficient vector. In (4), $\mathbf{T}$ is a highly rank deficient rectangular matrix that relates the spectral quantization of the hyperspectral imaging system to the high spatial resolution imaging system. Using the associativity between the matrices:

$$\mathbf{y} \approx \mathbf{T}(\boldsymbol{\Phi}\boldsymbol{\beta}) \approx \mathbf{T}\mathbf{s} \tag{5}$$

where, $\mathbf{s} \in \mathbb{R}^L$ denotes the pixel in the target image $\mathbf{S}$. Equation (5) suggests, if $\boldsymbol{\Phi}$ is known, the super-resolution hyperspectral image can be estimated using an appropriate coefficient matrix, without the need of inverting the highly rank deficient matrix $\mathbf{T}$.

4 Proposed solution

Let $\mathcal{D}$ be a finite collection of unit-norm vectors in $\mathbb{R}^L$. In our settings, $\mathcal{D}$ is the *dictionary* whose elements (i.e. the *atoms*) are denoted by $\boldsymbol{\varphi}_\omega$, where ω ranges

over an index set Ω. More precisely, $\mathcal{D} \stackrel{\text{def}}{=} \{\boldsymbol{\varphi}_\omega : \omega \in \Omega\} \subset \mathbb{R}^L$. Considering (3)-(5), we are interested in forming the matrix $\boldsymbol{\Phi}$ from $\mathcal{D}$, such that

$$\bar{\mathbf{Y}}_h \approx \boldsymbol{\Phi}\mathbf{A} \tag{6}$$

where, $\bar{\mathbf{Y}}_h \in \mathbb{R}^{L \times mn}$ is the matrix formed by concatenating the pixels of the hyperspectral image $\mathbf{Y}_h$ and $\mathbf{A}$ is the coefficient matrix with $\boldsymbol{\alpha}_i$ as its i^{th} column. We propose to draw $\boldsymbol{\Phi}$ from $\mathbb{R}^{L \times k}$, such that $k > q$; see (2). This is because, the LMM in (2) approximates a pixel assuming *linear* mixing of the material reflectances. In the real world, phenomena like multiple light scattering and existence of intimate material mixtures also cause non-linear mixing of the spectral signatures [26]. This usually alters the reflectance spectrum of a material or results in multiple distinct reflectance spectra of the same material in the scene. The matrix $\boldsymbol{\Phi}$ must also account for these spectra. Henceforth, we use the term *dictionary* for the matrix $\boldsymbol{\Phi}$[1].

According to the model in (6), each column of $\bar{\mathbf{Y}}_h$ is constructed using a very small number of dictionary atoms. Furthermore, the atoms of the dictionary are non-negative vectors as they correspond to reflectance spectra. Therefore, we propose to solve the following constrained sparse representation problem to learn the proposed dictionary $\boldsymbol{\Phi}$:

$$\min_{\Phi, A} ||\mathbf{A}||_1 \text{ s.t. } ||\bar{\mathbf{Y}}_h - \boldsymbol{\Phi}\mathbf{A}||_F \leq \eta, \boldsymbol{\varphi}_\omega \geq \mathbf{0}, \ \forall \omega \in \{1, ..., k\} \tag{7}$$

where, $||.||_1$ and $||.||_F$ denote the l_1 and the Forbenious norms of the matrices respectively, and η represents the modeling error. To solve (7) we use the online dictionary learning approach proposed by Mairal et al. [29] with an additional non-negativtiy constraint on the dictionary atoms - we refer the reader to the original work for details.

Once $\boldsymbol{\Phi}$ is known, we must compute an appropriate coefficient matrix $\mathbf{B} \in \mathbb{R}^{k \times MN}$; as suggested by (5), to estimate the target image $\mathbf{S}$. This matrix is computed using the learned dictionary and the image $\mathbf{Y}$ along with two important pieces of prior information. a) In the high spatial resolution image, nearby pixels are likely to represent the same materials in the scene. Hence, they should be well approximated by a small group of the same dictionary atoms. b) The elements of $\mathbf{B}$ must be non-negative quantities because they represent the fractional abundances of the spectral signal sources in the scene. It is worth mentioning that we could also use (b) for $\mathbf{A}$ in (7), however, there we were interested only in $\boldsymbol{\Phi}$. Therefore, a non-negativity constraint over $\mathbf{A}$ was unnecessary. Neglecting this constraint in (7) additionally provides computational advantages in solving the optimization problem.

Considering (a), we process the image $\mathbf{Y}$ in terms of small disjoint spatial patches for computing the coefficient matrix. We denote each of the image patch by $\mathbf{P} \in \mathbb{R}^{M_P \times N_P \times l}$ and estimate its corresponding coefficient matrix

[1] Formally, $\boldsymbol{\Phi}$ is the *dictionary synthesis matrix* [10]. However, we follow the convention of the previous literature in dictionary learning (e.g. [27], [28]), which rarely distinguishes the synthesis matrix from the dictionary.

$\mathbf{B}_P \in \mathbb{R}^{k \times M_P N_P}$ by solving the following constrained simultaneous sparse approximation problem:

$$\min_{B_P} ||\mathbf{B}_P||_{row\_0} \text{ s.t. } ||\bar{\mathbf{P}} - \widetilde{\boldsymbol{\Phi}}\mathbf{B}_P||_F \leq \varepsilon,\ \boldsymbol{\beta}_{p^i} \geq \mathbf{0}\ \forall i \in \{1, ..., M_P N_P\} \quad (8)$$

where, $\bar{\mathbf{P}} \in \mathbb{R}^{l \times M_p N_p}$ is formed by concatenating the pixels in $\mathbf{P}$, $\widetilde{\boldsymbol{\Phi}} \in \mathbb{R}^{l \times k}$ is the transformed dictionary i.e. $\widetilde{\boldsymbol{\Phi}} = \mathbf{T}\boldsymbol{\Phi}$; see (4), and $\boldsymbol{\beta}_{p^i}$ denotes the i^{th} column of the matrix $\mathbf{B}_p$. In the above objective function, $||.||_{row\_0}$ denotes the row-l_0 quasi-norm [10] of the matrix, which represents the cardinality of its row-support[2]. Formally,

$$||\mathbf{B}_p||_{row\_0} \stackrel{\text{def}}{=} \left| \bigcup_{i=1}^{M_p N_p} \text{supp}(\boldsymbol{\beta}_{p^i}) \right|$$

where, supp(.) indicates the support of a vector and |.| denotes the cardinality of a set. Tropp [28] has argued that (8) is an NP-hard problem without the non-negativity constraint. The combinatorial complexity of the problem does not change with the non-negativity constraint over the coefficient matrix. Therefore, the problem must either be relaxed [28] or solved by the greedy pursuit strategy [10]. We prefer the latter because of its computational advantages [30] and propose a simultaneous greedy pursuit algorithm, called G-SOMP+, for solving (8). The proposed algorithm is a generalization of a popular greedy pursuit algorithm Simultaneous Orthogonal Matching Pursuit (SOMP) [10], which additionally constrains the solution space to non-negative matrices. Hence, we denote it as G-SOMP+. Here, the notion of 'generalization' is similar to the one used in [31] that allows selection of multiple dictionary atoms in each iteration of Orthogonal Matching Pursuit (OMP) [32] to generalize OMP.

G-SOMP+ is given below as Algorithm 1. The algorithm seeks an approximation of the input matrix $\bar{\mathbf{P}}$ - henceforth, called the *patch* - by selecting the dictionary atoms $\tilde{\varphi}_\xi$ indexed in a set $\Xi \subset \Omega$, such that, $|\Xi| \ll |\Omega|$ and every $\tilde{\varphi}_\xi$ contributes to the approximation of the *whole* patch. In its i^{th} iteration, the algorithm first computes the cumulative correlation of each dictionary atom with the residue of its current approximation of the patch (line 5 in Algorithm 1) - the patch itself is considered as the residue for initialization. Then, it identifies L (an algorithm parameter) dictionary atoms with the highest cumulative correlations. These atoms are added to a subspace indexed in a set Ξ^i, which is empty at initialization. The aforementioned subspace is then used for a non-negative least squares approximation of the patch (line 8 in Algorithm 1) and the residue is updated. The algorithm stops if the updated residue is more than a fraction γ of the residue in the previous iteration. Note that, the elements of the set Ξ in G-SOMP+ also denote the row-support of the coefficient matrix. This is because, a dictionary atom can only participate in the patch approximation if the corresponding row of the coefficient matrix has some non-zero element in it.

[2] Set of indices for the non-zero rows of the matrix.

Algorithm 1 G-SOMP+

Initializaiton:

1: Iteration: $i = 0$
2: Initial solution: $\mathbf{B}^0 = \mathbf{0}$
3: Initial residue: $\mathbf{R}^0 = \bar{\mathbf{P}} - \widetilde{\boldsymbol{\Phi}}\mathbf{B}^0 = \bar{\mathbf{P}}$
4: Initial index set: $\Xi^0 = \emptyset$ = row-supp$\{\mathbf{B}^0\}$, row-supp$\{\mathbf{B}\} = \{1 \leq t \leq k : \boldsymbol{\beta}^t \neq \mathbf{0}\}$, where $\boldsymbol{\beta}^t$ is the t^{th} row of $\mathbf{B}$.

Main Iteration: Update iteration: $i = i + 1$

5: Compute $b_j = \sum_{\tau=1}^{M_P N_P} \frac{\widetilde{\boldsymbol{\Phi}}_j^T \mathbf{R}_\tau^{i-1}}{||\mathbf{R}_\tau^{i-1}||_2^2}$, $\forall j \in \{1, ..., k\}$, where, $\mathbf{X}_z$ denotes the z^{th} column of the matrix $\mathbf{X}$.
6: $\mathcal{N} = \{$indices of $\widetilde{\boldsymbol{\Phi}}$'s atoms corresponding to the L largest $b_j\}$
7: $\Xi^i = \Xi^{i-1} \bigcup \mathcal{N}$
8: $\mathbf{B}^i = \min ||\widetilde{\boldsymbol{\Phi}}\mathbf{B} - \bar{\mathbf{P}}||_F^2$ s.t. row-supp$\{\mathbf{B}\} = \Xi^i$, $\boldsymbol{\beta}^t \geq \mathbf{0}, \forall t$
9: $\mathbf{R}^i = \bar{\mathbf{P}} - \widetilde{\boldsymbol{\Phi}}\mathbf{B}^i$
10: If $||\mathbf{R}^i||_F > \gamma ||\mathbf{R}^{i-1}||_F$ stop, otherwise iterate again.

G-SOMP+ has three major differences from SOMP. 1) Instead of integrating the *absolute* correlations, it sums the correlations between a dictionary atom and the residue vectors (line 5 of Algorithm 1). 2) It approximates the patch in each iteration with the *non-negative* least squares method, instead of using the standard least squares approximation. 3) It selects L dictionary atoms in each iteration instead of a single dictionary atom. In the above mentioned difference, (1) and (2) impose the non-negativety constraint over the desired coefficient matrix. On the other hand, (3) primarily aims at improving the computation time of the algorithm. G-SOMP+ also uses a different stopping criterion than SOMP, that is controlled by γ - the residual decay parameter. We defer further discussion on (3) and the stopping criterion to Section 6. G-SOMP+ has been proposed specifically to solve the constrained simultaneous sparse approximation problem in (8). Therefore, it is able to approximate a patch better than a generic greedy pursuit algorithm (e.g. SOMP).

Solving (8) for each image patch results in the desired coefficient matrix $\mathbf{B}$ that is used with $\boldsymbol{\Phi}$ to compute $\hat{\bar{\mathbf{S}}} \in \mathbb{R}^{L \times MN}$, which is the estimate of the super-resolution hyperspectral image $\bar{\mathbf{S}} \in \mathbb{R}^{L \times MN}$ (in matrix form).

$$\hat{\bar{\mathbf{S}}} = \boldsymbol{\Phi}\mathbf{B} \tag{9}$$

Fig. 1 pictorially summarizes the proposed approach.

5 Experimental results

We have evaluated our approach using ground based hyperspectral images as well as remotely sensed data. For the ground based images, we have conducted experiments with two different public databases. The first database [33], called the CAVE database, consists of 32 hyperspectral images of everyday objects. The

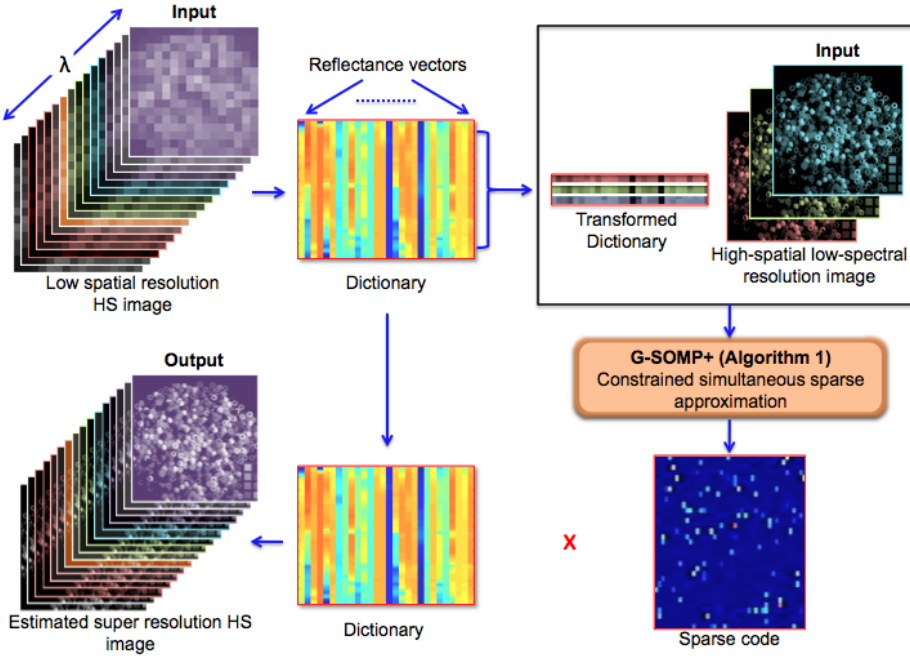


Fig. 1: Schematic of the proposed approach: The low spatial resolution hyperspectral (HS) image is used for learning a dictionary whose atoms represent reflectance spectra. This dictionary is transformed and used with the high-spatial but low-spectral resolution image to learn a sparse code by solving a constrained simultaneous sparse approximation problem. The sparse code is used with the original dictionary to estimate the super-resolution HS image.

512×512 spectral images of the scenes are acquired at a wavelength interval of 10 nm in the range $400-700$ nm. The second is the Harvard database [34], which consists of hyperspectral images of 50 real-world indoor and outdoor scenes. The 1392×1040 spectral images are sampled at every 10 nm from 420 to 720 nm. Hyperspectral images of the databases are considered as the ground truth for the super-resolution hyperspectral images. We down-sample a ground truth image by averaging over 32×32 disjoint spatial blocks to simulate the low spatial resolution hyperspectral image $\mathbf{Y}_h$. From the Harvard database, we have only used 1024×1024 image patches to match the down-sampling strategy. Following [22], a high spatial (but low spectral) resolution image $\mathbf{Y}$ is created by integrating a ground truth image over the spectral dimension, using the Nikon D700 spectral response[3] - which makes $\mathbf{Y}$ a simulated RGB image of the same scene. Here, we present the results on eight representative images from each database, shown in Fig. 2. We have selected these images based on the variety of the scenes. Results on further images are provided in the supplementary material of the paper.

[3] `https://www.maxmax.com/spectral_response.htm`

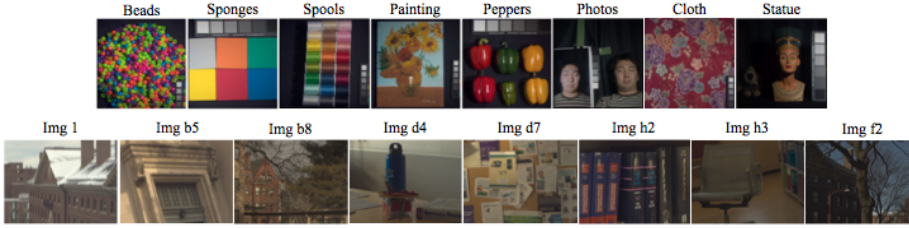


Fig. 2: RGB images from the databases. First row: Images from the CAVE database [33]. Second row: Images from the Harvard database [34].

Fig. 3 shows the results of using our approach for estimating the super-resolution hyperspectral images of 'Painting' and 'Peppers' (see Fig. 2). The top row shows the input 16×16 hyperspectral images at 460, 540 and 620 nm. The ground truth images at these wavelengths are shown in the second row, which are clearly well approximated in the estimated images shown in the third row. The fourth row of the figure shows the difference between the ground truth images and the estimated images. The results demonstrate a successful estimation of the super-resolution spectral images. Following the protocol of [3] and [22], we have used Root Mean Square Error (RMSE) as the metric for further quantitative evaluation of the proposed approach and its comparison with the existing methods.

$$RMSE = \sqrt{\frac{||\bar{\mathbf{S}} - \hat{\bar{\mathbf{S}}}||_{\mathbf{F}}^{\mathbf{2}}}{LMN}} \tag{10}$$

where, $\bar{\mathbf{S}}$ and $\hat{\bar{\mathbf{S}}}$ respectively denote the ground truth image and the estimated image as matrices in $\mathbb{R}^{L \times MN}$. Table 1 shows the RMSE values of the proposed approach and the existing methods for the images of the CAVE database [33]. Among the existing approaches we have chosen the Matrix Factorization method (MF) in [3], the Spatial and Spectral Fusion Model (SASFM) [8], the ADMM based method [22] and the Coupled Matrix Factorization method (CMF) [23] for the comparison. Most of these matrix factorization based approaches have been shown to outperform the other techniques discussed in Section 2. To show the difference in the performance, Table 1 also includes some results from the Component Substitution Method (CSM) [14] - taken directly from [3]. We have used our own implementations of MF and SASFM because of unavailability of the public codes from the authors. To ensure an un-biased comparison, we take special care that the results achieved by our implementations are either the same or better than the results reported originally by the authors on the same images. Needless to mention, we follow the same experimental protocol as the previous works. The results of CSM and ADMM are taken directly form [22]. Note that, these algorithms also require *a priori* knowledge of the spatial transform between the hyperspectral image and the high resolution image, because of which they

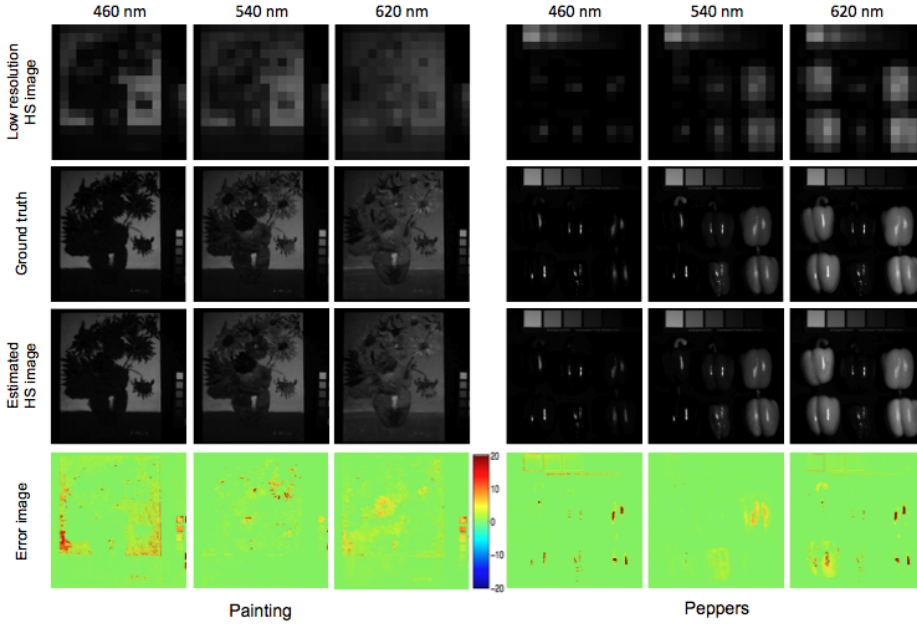


Fig. 3: Spectral images for Painting (Left) and Peppers (Right) at 460, 540 and 620 nm. Top row: 16×16 low spatial resolution hyperspectral (HS) images. Second row: 512×512 ground truth images. Third row: Estimated 512×512 HS images. Fourth row: Corresponding error images, where the scale is in the range of 8 bit images.

are highlighted in red in the table. The proposed approach has been able to outperform these methods without requiring this knowledge.

For the proposed approach, we have used 75 atoms in the dictionary and let $L = 20$ for each iteration of G-SOMP+, which processes 8×8 image patches. We have chosen $\eta = 10^{-5}$ in (7) and the residual decay parameter of G-SOMP+, $\gamma = 0.99$. We have optimized these parameter values, and the parameter settings of MF and SASFM, using a separate training set of 30 images. The training set comprises 15 images selected at random from each of the used databases. We have used the same parameter settings for all the results reported here and in the supplementary material. We defer further discussion on the parameter value selection for the proposed approach to Section 6.

Results on the images from the Harvard database [34] are shown in Table 2. In this table, we have compared the results of the proposed approach only with MF and SASFM because, like our approach, only these two approaches do not require the knowledge of the spatial transform between the input images. The table shows that the proposed approach consistently performs better than others. We have also experimented with the hyperspectral data that is remotely sensed by the NASA's Airborne Visible and Infrared Imaging Spectrometer (AVIRIS) [35].

Table 1: Comparison of the approaches using [33]. The reported RMSE values are in the range of 8 bit images. The best results are shown in bold. The approaches highlighted in red also require the knowledge of spatial transform between the input images, which restrict their practical applicability.

	CAVE database [33]							
Method	Beads	Sponges	Spools	Painting	Pepper	Photos	Cloth	Statue
CSM [14]	28.5	19.9	-	12.2	13.7	13.1	-	-
MF [3]	8.2	3.7	8.4	4.4	4.6	3.3	6.1	2.7
SASFM [8]	9.2	5.3	6.1	4.3	6.3	3.7	10.2	3.3
ADMM [22]	6.1	2.0	5.3	6.7	2.1	3.4	9.5	4.3
CMF [23]	6.6	4.0	15.0	26.0	5.5	11.0	20.0	16.0
Proposed	**3.7**	**1.5**	**3.8**	**1.3**	**1.3**	**1.8**	**2.4**	**0.6**

Table 2: Comparison of the approaches using [34]. The reported RMSE values are in the range of 8 bit images. The best results are shown in bold.

	Harvard database [34]							
Method	Img 1	Img b5	Img b8	Img d4	Img d7	Img h2	Img h3	Img f2
MF [3]	3.9	2.8	6.9	3.6	3.9	3.7	2.1	3.1
SASFM [8]	4.3	2.6	7.6	4.0	4.0	4.1	2.3	2.9
Proposed	**1.2**	**0.9**	**2.8**	**0.8**	**1.2**	**1.6**	**0.5**	**0.9**

AVIRIS samples the scene reflectance in the wavelength range 400 - 2500 nm at a nominal interval of 10 nm. We have used a hyperspectral image taken over the Cuprite mines, Nevada[4]. The image has dimensions $512 \times 512 \times 224$, where 224 represents the number of spectral bands in the image. Following [26], we have removed the bands 1-2, 105-115, 150-170 and 223-224 of the image because of extremely low SNR and water absorptions in those bands. We perform the down-sampling on the image as before and construct $\mathbf{Y}$ by directly selecting the 512×512 spectral images from the ground truth image, corresponding to the wavelengths 480, 560, 660, 830, 1650 and 2220 nm. These wavelengths correspond to the visible and mid-infrared range spectral channels of USGS/NASA Landsat 7 satellite[5]. We adopt this strategy of constructing $\mathbf{Y}$ from Huang et al. [8]. Fig. 4 shows the results of our approach for the estimation of the super-resolution hyperspectral image at 460, 540, 620 and 1300 nm. For this data set, the RMSE values for the proposed approach, MF [3] and SASFM [8] are 1.12, 3.06 and 3.11, respectively.

[4] Available at `http://aviris.jpl.nasa.gov/data/free_data.html`.
[5] `http://www.satimagingcorp.com/satellite-sensors/landsat.html`.

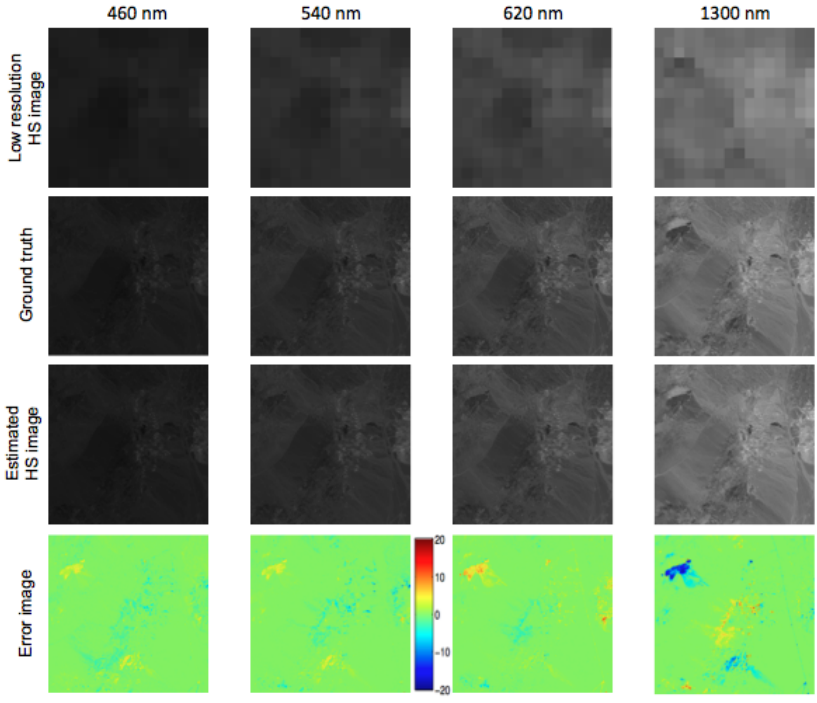
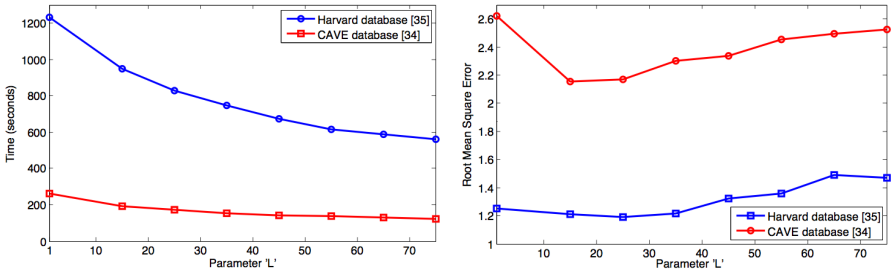


Fig. 4: Spectral images for AVIRIS data at 460, 540, 620 and 1300 nm. Top row: 16×16 low spatial resolution hyperspectral (HS) image. Second row: 512×512 ground truth image. Third row: Estimated 512×512 HS image. Fourth row: Corresponding error image, with the scale is in the range of 8 bit images.

6 Discussion

G-SOMP+ uses two parameters. L: the number of dictionary atoms selected in each iteration, and γ: the residual decay parameter. By selecting more dictionary atoms in each iteration, G-SOMP+ computes the solution more quickly. The processing time of G-SOMP+ as a function of L, is shown in Fig. 5a. Each curve in Fig. 5 represents the mean values computed over a separate training data set of 15 images randomly selected from the database, whereas the dictionary used by G-SOMP+ contained 75 atoms. Fig. 5a shows the timings on an Intel Core i7-2600 CPU at 3.4 GHz with 8 GB RAM. Fig. 5b shows the RMSE values on the training data set as a function of L. Although, the error is fairly small over the complete range of L, the values are particularly low for $L \in \{15, ..., 25\}$, for both of the databases. Therefore, we have chosen $L = 20$ for all the test images in our experiments. Incidentally, the number of distinct spectral sources in a typical remote sensing hyperspectral image is also considered to be close to 20 [25]. Therefore, we have used the same value of the parameter for the remote sensing test image.



(a) (b)

Fig. 5: Selection of the G-SOMP+ parameter L: The values are the means computed over 15 separate training images for each database: a) Processing time of G-SOMP+ in seconds as a function of L. The values are computed on an Intel Core i7-2600 CPU at 3.4 GHz with 8 GB RAM. b) RMSE of the estimated images by G-SOMP+ as a function of L.

Generally, it is hard to know *a priori* the exact number of iterations required by a greedy pursuit algorithm to converge. Similarly, if the residual error (i.e. $||\mathbf{R}^i||_F$ in Algorithm 1) is used as the stopping criterion, it is often difficult to select a single best value of this parameter for all the images. Fig. 5b shows that the RMSE curves rise for the higher values of L after touching a minimum value. In other words, more than the required number of dictionary atoms adversely affect the signal approximation. We use this observation to decide on the stopping criterion of G-SOMP+. Since the algorithm selects a constant number of atoms in each iteration, it stops if the approximation residual in its current iteration is more than a fraction γ of the residual in the previous iteration. As the approximation residual generally decreases rapidly before increasing (or becoming constant in some cases), we found that the performance of G-SOMP+ on the training images was mainly insensitive for $\gamma \in [0.75, 1]$. From this range, we have selected $\gamma = 0.99$ for the test images in our experiments.

Our approach uses the online-dictionary learning technique [29] to solve (7). This technique needs to know the total number of dictionary atoms to be learned *a priori*. In Section 4, we have argued to use more dictionary atoms than the number of distinct materials in the scene. This results in a better separation of the spectral signal sources in the scene. Fig. 6 illustrates this notion. The figure shows an RGB image of 'Sponges' on the left. To extract the reflectance spectra, we learn two different dictionaries with 10 and 50 atoms, respectively, using the 16×16 hyperspectral image of the scene. We cluster the atoms of these dictionaries based on their correlation and show the arranged dictionaries in Fig. 6. From the figure, we can see that the dictionary with 10 atoms is not able to clearly distinguish between the reflectance spectra of the blue (C1) and the green (C2) sponge, whereas 10 seems to be a reasonable number representing the distinct materials in the scene. On the other hand, the dictionary with 50 atoms has learned two separate clusters for the two sponges.

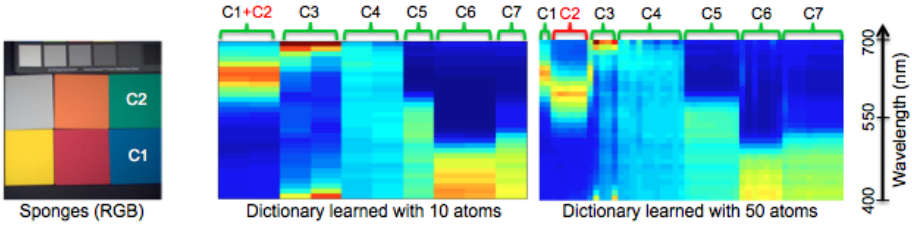


Fig. 6: Selecting the number of dictionary atoms: RGB image of 'Sponges', containing roughly $7-10$ distinct colors (materials), is shown on the left. Two dictionaries, with 10 and 50 atoms, are learned for the scene. After clustering the spectra (i.e. the dictionary atoms) into seven clusters (C1 - C7), it is visible that the dictionary with 50 atoms learns distinct clusters for the blue (C1) and the green (C2) sponges, whereas the dictionary with 10 atoms is not able to clearly distinguish between these sponges.

The results reported in Fig. 5 are relatively insensitive to the number of dictionary atoms in the range of 50 to 80. In all our experiments, the proposed approach has learned a dictionary with 75 atoms. We choose a larger number to further incorporate the spectral variability of highly mixed scenes.

7 Conclusion

We have proposed a sparse representation based approach for hyperspectral image super-resoltuion. The proposed approach fuses a high spatial (but low spectral) resolution image with the hyperspectral image of the same scene. It uses the input low resolution hyperspectral image to learn a dictionary by solving a constrained sparse optimization problem. The atoms of the learned dictionary represent the reflectance spectra related to the scene. The learned dictionary is transformed according to the spectral quantization of the input high resolution image. This image and the transformed dictionary are later employed by an algorithm G-SOMP+. The proposed algorithm efficiently solves a constrained simultaneous sparse approximation problem to learn a sparse code. This sparse code is used with the originally learned dictionary to estimate the super-resolution hyperspectral image of the scene. We have tested our approach using the hyperspectral images of objects, real-world indoor and outdoor scenes and a remotely sensed hyeprspectral image. Results of the experiments demonstrate that by taking advantage of the signal sparsity, non-negativity and the spatial structure in the scene, the proposed approach is able to consistently perform better than the existing state of the art methods on all the data sets.

8 Acknowledgements

This research was supported by ARC Discovery Grant DP110102399.

References

1. Bioucas-Dias, J., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., Chanussot, J.: Hyperspectral remote sensing data analysis and future challenges. IEEE Geosci. Remote Sens. Mag. **1** (2013) 6–36
2. Akhtar, N., Shafait, F., Mian, A.: Repeated constrained sparse coding with partial dictionaries for hyperspectral unmixing. In: IEEE Winter Conference on Applications of Computer Vision. (2014)
3. Kawakami, R., Wright, J., Tai, Y.W., Matsushita, Y., Ben-Ezra, M., Ikeuchi, K.: High-resolution hyperspectral imaging via matrix factorization. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). (2011) 2329–2336
4. Nguyen, H.V., Banerjee, A., Chellappa, R.: Tracking via object reflectance using a hyperspectral video camera. In: IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW). (2010) 44 – 51
5. Tarabalka, Y., Chanussot, J., Benediktsson, J.A.: Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers. IEEE Trans. Syst., Man, Cybern., Syst. **40**(5) (2010) 1267–1279
6. Uzair, M., Mahmood, A., Mian, A.: Hyperspectral face recognition using 3D-DCT and partial least squares. In: British Machine Vision Conf. (BMVC). (2013)
7. Khan, Z., Shafait, F., Mian, A.: Hyperspectral imaging for ink mismatch detection. In: Int. Conf. on Document Analysis and Recognition (ICDAR). (2013)
8. Huang, B., Song, H., Cui, H., Peng, J., Xu, Z.: Spatial and spectral image fusion using sparse matrix factorization. IEEE Trans. Geosci. Remote Sens. **52**(3) (March 2014) 1693–1704
9. Olshausen, B.A., Fieldt, D.J.: Sparse coding with an overcomplete basis set: a strategy employed by v1. Vision Research **37** (1997) 3311–3325
10. Tropp, J.A., Gilbert, A.C., Strauss, M.J.: Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. Signal Processing **86**(3) (2006) 572–588
11. Wang, Z., Ziou, D., Armenakis, C., Li, D., Li, Q.: A comparative analysis of image fusion methods. IEEE Trans. Geosci. Remote Sens. **43**(6) (June 2005) 1391–1402
12. Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P., Bruce, L.: Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest. IEEE Trans. Geosci. Remote Sens. **45**(10) (Oct 2007) 3012–3021
13. Carper, W.J., Lilles, T.M., Kiefer, R.W.: The use of intensity-hue-saturation transformations for merging SOPT panchromatic and multispectrl image data. Photogram. Eng. Remote Sens. **56**(4) (1990) 459 – 467
14. Aiazzi, B., Baronti, S., Selva, M.: Improving component substitution pansharpening through multivariate regression of MS +Pan data. IEEE Trans. Geosci. Remote Sens. **45**(10) (Oct 2007) 3230–3239
15. Imai, F.H., Berns, R.S.: High resolution multispectral image archives: a hybrid approach. In: Color Imaging Conference. (1998) 224 – 227
16. Koutsias, N., Karteris, M., Chuvieco, E.: The use of intensity hue saturation transformation of Landsat 5 Thematic Mapper data for burned land mapping. Photogram. Eng. Remote Sens. **66**(7) (2000) 829 – 839
17. Cetin, M., Musaoglu, N.: Merging hyperspectral and panchromatic image data: Qualitative and quantitative analysis. Int. J. Remote Sens. **30**(7) (January 2009) 1779–1804
18. Minghelli-Roman, A., Polidori, L., Mathieu-Blanc, S., Loubersac, L., Cauneau, F.: Spatial resolution improvement by merging MERIS-ETM images for coastal water monitoring. IEEE Geosci. Remote Sens. Lett. **3**(2) (April 2006) 227–231

19. Zhukov, B., Oertel, D., Lanzl, F., Reinhackel, G.: Unmixing-based multisensor multiresolution image fusion. IEEE Trans. Geosci. Remote Sens. **37**(3) (May 1999) 1212–1226
20. Zurita-Milla, R., Clevers, J.G., Schaepman, M.E.: Unmixing-based Landsat TM and MERIS FR data fusion. IEEE Trans. Geosci. Remote Sens. **5**(3) (2008) 453–457
21. Kopf, J., Cohen, M.F., Lischinski, D., Uyttendaele, M.: Joint bilateral upsampling. ACM Trans. Graph. **26**(3) (July 2007)
22. Wycoff, E., Chan, T.H., Jia, K., Ma, W.K., Ma, Y.: A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. In: IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP). (2013)
23. Yokoya, N., Yairi, T., Iwasaki, A.: Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. IEEE Trans. Geosci. Remote Sens. **50**(2) (Feb 2012) 528–537
24. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1) (January 2011) 1–122
25. Keshava, N., Mustard, J.: Spectral unmixing. IEEE Signal Process. Mag. **19**(1) (Jan 2002) 44–57
26. Iordache, M.D., Bioucas-Dias, J., Plaza, A.: Sparse unmixing of hyperspectral data. IEEE Trans. Geosci. Remote Sens. **49**(6) (June 2011) 2014–2039
27. Aharon, M., Elad, M., Bruckstein, A.: K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. **54**(11) (Nov 2006) 4311–4322
28. Tropp, J.A.: Algorithms for simultaneous sparse approximation. part ii: Convex relaxation. Signal Processing **86**(3) (2006) 589 – 602
29. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: Int. Conf. on Machine Learning. ICML '09 (2009) 689–696
30. Bruckstein, A., Elad, M., Zibulevsky, M.: On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations. IEEE Trans. Inf. Theory **54**(11) (Nov 2008) 4813–4820
31. Wang, J., Kwon, S., Shim, B.: Generalized orthogonal matching pursuit. IEEE Trans. Signal Process. **60**(12) (Dec 2012) 6202–6216
32. Tropp, J., Gilbert, A.: Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Inf. Theory **53**(12) (Dec 2007) 4655–4666
33. Yasuma, F., Mitsunaga, T., Iso, D., Nayar, S.: Generalized assorted pixel camera: Post-capture control of resolution, dynamic range and spectrum. Technical report (Nov 2008)
34. Chakrabarti, A., Zickler, T.: Statistics of real-world hyperspectral images. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). (2011) 193–200
35. Green, R.O., Eastwood, M.L., Sarture, C.M., Chrien, T.G., Aronsson, M., Chippendale, B.J., Faust, J.A., Pavri, B.E., Chovit, C.J., Solis, M., Olah, M.R., Williams, O.: Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). Remote Sensing of Environment **65**(3) (1998) 227 – 248