Compressive Sensing and Adaptive Direct Sampling in Hyperspectral Imaging

Jürgen Hahn[a], Christian Debes[b], Michael Leigsnering[a], Abdelhak M. Zoubir[a]

[a]*Signal Processing Group, Institute of Telecommunications, Technische Universität Darmstadt, Merckstr. 25, 64283 Darmstadt, Germany*
[b]*AGT Group (R&D) GmbH, Hilpertstr. 20a, 64295 Darmstadt, Germany*

Abstract

Hyperspectral imaging (HSI) is an emerging technique, which provides the continuous acquisition of electro-magnetic waves, usually covering the visible as well as the infrared light range. Many materials can be easily discriminated by means of their spectra rendering HSI an interesting method for the reliable classification of contents in a scene. Due to the high amount of data generated by HSI, effective compression algorithms are required. The computational complexity as well as the potentially high number of sensors render HSI an expensive technology. It is thus of practical interest to reduce the number of required sensor elements as well as computational complexity – either for cost or for energy reasons. In this paper, we present two different systems that acquire hyperspectral images with less samples than the actual number of pixels, i.e. in a low dimensional representation. First, a design based on compressive sensing (CS) is explained. Second, adaptive direct sampling (ADS) is utilized to obtain coefficients of hyperspectral images in the 3D (Haar) wavelet domain, simplifying the reconstruction process significantly. Both approaches are compared with conventionally captured images with respect to image quality and classification accuracy. Our results based on real data show that in most cases only 40% of the samples suffice to obtain high quality images. Using ADS, the rate can be reduced even to a greater extent. Further results confirm that although the number of acquired samples is dramatically reduced, we can still obtain high classification rates.

Keywords: Compressive sensing, hyperspectral imaging, classification, image reconstruction, sampling strategy

Email addresses: hahn@spg.tu-darmstadt.de (Jürgen Hahn), cdebes@agtgermany.com (Christian Debes), leigsnering@spg.tu-darmstadt.de (Michael Leigsnering), zoubir@spg.tu-darmstadt.de (Abdelhak M. Zoubir)

1. Introduction

Hyperspectral imaging (HSI), frequently also described as imaging spectroscopy, is the fusion of spectroscopy and image processing. Thus, hyperspectral images contain the reflectance of the visible and infrared light decomposed in tens or hundreds of bands. It is a modern, emerging optical acquisition technique.

Objects of different materials absorb and reflect light at different frequencies due to their molecular structure [1], resulting in discriminative spectra or signatures. Therefore, identification of a great diversity of materials becomes fairly easy, rendering HSI interesting for a wide area of engineering tasks and research.

Today, HSI is often used in mineralogy, agriculture and surveillance [2]. In agriculture, HSI is mainly used for the examination of the condition and growth of crop to prevent shortages of comestible goods [3, 4]. There is an increasing interest in surveillance of large areas for security and defense purposes. Here, HSI plays an increasingly important role because it allows accurate detection of anomalies [5, 6]. For example, the problem of detecting buried landmines has been successfully performed with HSI [7, 8]. Thanks to further developments, more and more researchers from different fields explore the potential of HSI. Especially, industry has just started to realize the full capability of HSI, for example in the field of nutrition analysis [9] and waste recycling [10].

Especially for mobile systems, e.g. aircrafts, satellites or unmanned aerial vehicles (UAV), where the data is usually compressed on board to save transmission time and bandwidth, computation power and energy are strongly limited resources. Opposed to this, compression is computationally expensive and stresses resources. However, omitting compression rarely presents a solution as the required bandwidth would dramatically increase. Consequently, the question arises whether it is possible to capture a low dimensional representation of the image directly. This question has been recently answered by compressive sensing (CS) [11, 12], allowing for considerable undersampling of the signal by using reconstruction algorithms after the acquisition. CS has been employed in many different fields, such as magnetic resonance imaging (MRI) [13] and radar imaging [14, 15].

In the recent past, a few papers have been published that were concerned with an adaption of compressive sensing to HSI. Golbabaee *et al.* [16] perform CS on the whole hyperspectral image cube. A drawback of their method is the utilization of a least-squares approach on the mobile system to reduce the number of spectral bands. Also, a library of spectral signatures is required, which includes all information contained in the image. In remote sensing, it is rarely known which information is in the scene and thus, this requirement poses a strong limitation on the system proposed in [16]. For further compression, a thresholding scheme is applied to the least-squares method to decide which information is present in the scene. This can be considered as a classification performed in the mobile system, though more sophisticated methods exist for this purpose. Since the least-squares method is sensitive to (non-Gaussian)

noise, this method performs poorly on real images. Furthermore, the least squares step actually implies a classification that shows weak performance for real images.

Li *et al.* [17] additionally perform an unmixing scheme of the captured signatures. Instead of the least-squares method, a singular value decomposition of the observation matrix is performed that avoids the reconstruction of the full hyperspectral image cube and directly unmixes the data. The experimental results based on real data show a reduction of the number of measurements while maintaining a high image quality. Similar to Golbabaee *et al.* [16], Li *et al.* require prior knowledge of the scene in form of a dictionary of the endmember signatures, leading to the previously mentioned restrictions.

The Coded Aperture Snapshot Spectral Imaging system (CASSI) [18, 19] constitutes one of the most advanced developments. Herein, a coded aperture is used that allows capturing the scene by a single snapshot, exhibiting a great advantage over other systems. A prototype system demonstrated that high compression rates are achieved. However, the question arises whether this system is able to capture hundreds of bands as expected in hyperspectral imaging. The problem relies on the fact that both the spatial and spectral information are captured at once, i.e. a trade-off between spatial and spectral resolution has to be found. Apart from this, the sensor array given by a microchip must be able to capture the desired spectral range at any position. Especially, if visible light as well as infrared light is captured, it is a challenging task to design such a device.

Another sophisticated design is developed in [20]. Studer *et al.* present a microscopy system that is based on fluorescence microscopy. This system is also able to compressively capture hyperspectral images of 128 bands, ranging from ultraviolet to infrared. Studer *et al.* suggest the use of a Hadamard matrix as sensing matrix, which is modified so that it can easily be mapped onto a DMD. The results indicate high quality reconstruction, even for very low subsampling rates.

The contribution of this paper is a new framework for the joint acquisition and compression of hyperspectral images. Our aim is to provide conceptual proofs on how to capture hyperspectral images, in which the number of sensors or measurements is lower than the actual resolution of the image. We show how the concepts of (1) compressive sensing and (2) adaptive direct sampling can be applied to hyperspectral imaging, using (3) a novel sampling scheme for the latter. A strong reduction in terms of the number of measurements is achieved while maintaining a high classification rate. We believe that this allows for reduced optical and electronic hardware costs as well as lower computational load for mobile imaging systems, where the reconstruction of the image is performed offline at a base station. The performance of the proposed framework is assessed in terms of image quality as well as classification performance on real images.

This paper is structured as follows. In the next section, the fundamental basics of CS are introduced. In Section 3 we revisit the hyperspectral image model suggested by Golbabaee *et al.* [16] and use CS techniques to capture the image. In Section 4, we detail a novel adaptive direct sampling method

for HSI. Classification results of hyperspectral images, reconstructed by the proposed methods, are shown in Section 5 and discussed in Section 6. Finally, conclusions are drawn in Section 7.

2. Compressive Sensing

The key idea of compressive sensing (CS) [11, 12] is to capture or sense the signal of interest in a low dimensional representation. For this purpose, the signal is required to be sparse in a certain domain. Let $\mathbf{x} = (x_1, \ldots, x_N)^T$ denote a signal that can be sparsely represented by means of the transform matrix $\mathbf{\Psi}$ which describes an $N \times N$ orthonormal basis, so that

$$\mathbf{x} = \mathbf{\Psi^T s} \tag{1}$$

where $\mathbf{s} = (s_1, \ldots, s_N)^T$ denotes the K-sparse coefficient vector. The K-sparse property describes that only $K \ll N$ coefficients are non-zero.

In contrast to conventional sensing, following the *Whittaker-Kotelnikov-Shannon* theorem [21], not all N samples of the signal $\mathbf{x}$ are acquired in CS. The observed vector $\mathbf{y} = (y_1, \ldots, y_M)^T$ is captured using only $M < N$ measurements, i.e.

$$\mathbf{y} = \mathbf{\Phi x} = \mathbf{\Phi \Psi^T s} = \mathbf{A s} \tag{2}$$

where $\mathbf{\Phi} = [\boldsymbol{\varphi}_1 \ldots \boldsymbol{\varphi}_M]^T$ denotes the $M \times N$ measurement matrix. In general, a good choice of $\mathbf{\Phi}$ is given by a Gaussian or Bernoulli random matrix [12, 22]. Since we are interested in $\mathbf{x}$, an under-determined system needs to be solved, which is possible thanks to the sparsity of $\mathbf{s}$. Consequently, only M with $K \leq M \ll N$ measurements are actually required for perfect reconstruction [23]. Thus, the number of samples can be dramatically reduced.

Mathematically, the sparsity of $\mathbf{s}$ is exploited by regularization. Ideally, the l_0-norm would be employed, resulting however in an infeasible NP-hard problem. Donoho and Candès *et al.* [11, 12] showed that minimizing the l_1-norm produces comparable results, i.e. the reconstructed signal $\hat{\mathbf{s}}$ is obtained by

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad \text{s.t. } \mathbf{\Phi \Psi^T s} = \mathbf{y}. \tag{3}$$

For reconstructing images, Candès *et al.* [24] showed that the utilization of the total variation (TV) norm often leads to more accurate reconstructions than the l_1-norm with respect to the image quality. The reconstruction formulation is then given by

$$\hat{\mathbf{X}}^{(2\mathrm{D})} = \arg\min_{\mathbf{X}^{(2\mathrm{D})}} \|\mathbf{X}^{(2\mathrm{D})}\|_{\mathrm{TV}} \quad \text{s.t. } \mathbf{\Phi}\,\mathrm{vec}(\mathbf{X}^{(2\mathrm{D})}) = \mathbf{y}. \tag{4}$$

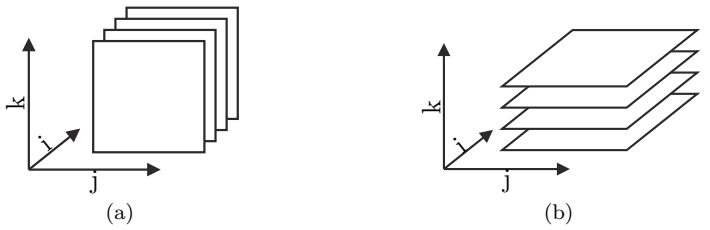


Figure 1: Decomposition of the hyperspectral image cube by means of (a) spectral stacking and (b) spatial stacking

where $\mathbf{X}^{(2D)} \in \mathbb{R}^{N_1 \times N_2}$ is the image and the TV norm is defined as

$$\|\mathbf{X}^{(2D)}\|_{TV} = \sum_{i}^{N_1} \sum_{j}^{N_2} \|(\nabla \mathbf{X}^{(2D)})_{i,j}\| \tag{5}$$

with $X_{i,j}^{(2D)}$ representing the (i,j)-th pixel value. In this equation, it becomes apparent that the TV-norm is closely related to minimizing the gradient of the image which is assumed to be sparse. Thus, a transform into a sparse domain is not required in this case.

3. Unstructured Measurement Matrices for HSI

In this section, we present different ways to capture hyperspectral images, using compressive sensing. Golbabaee *et al.* [16] considered the hyperspectral image as a stack of bands, i.e. gray value images. Thus, they basically performed CS on each band individually, which we refer to as spatial stacking. Alternatively, the hyperspectral image can also be decomposed into spatial-spectral layers denoted by spectral stacking. This approach is closely related to conventional hyperspectral imaging systems. Using these decompositions, CS of HSI is accomplished similarly to compressive imaging by means of a single-pixel camera [25, 26]. Both models of the hyperspectral image are depicted in Fig. 1.

3.1. Spectral stacking

For spectral stacking, we propose the system design shown in Fig. 2. Compared to the conventional *pushbroom* hyperspectral imaging system [1], the major modification consists in the replacement of the sensor array by a digital micromirror device (DMD) array whose reflected light is focused by a collection lens onto a photo diode, yielding an observation vector $\mathbf{y}_i$ for each layer $1 \leq i \leq N_1$. In the sequel, the hyperspectral image is denoted by $\mathbf{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ where (N_1, N_2) and N_3 are the spatial and the spectral resolution, respectively. The measurement matrix $\mathbf{\Phi}$ represented by the DMD array in Fig. 2 is a Bernoulli matrix of size $M_{23} \times N_{23}$ with $M_{23} \ll N_{23} = N_2 N_3$ denoting the number of measurements for each layer, since the DMD can only represent positive binary values.

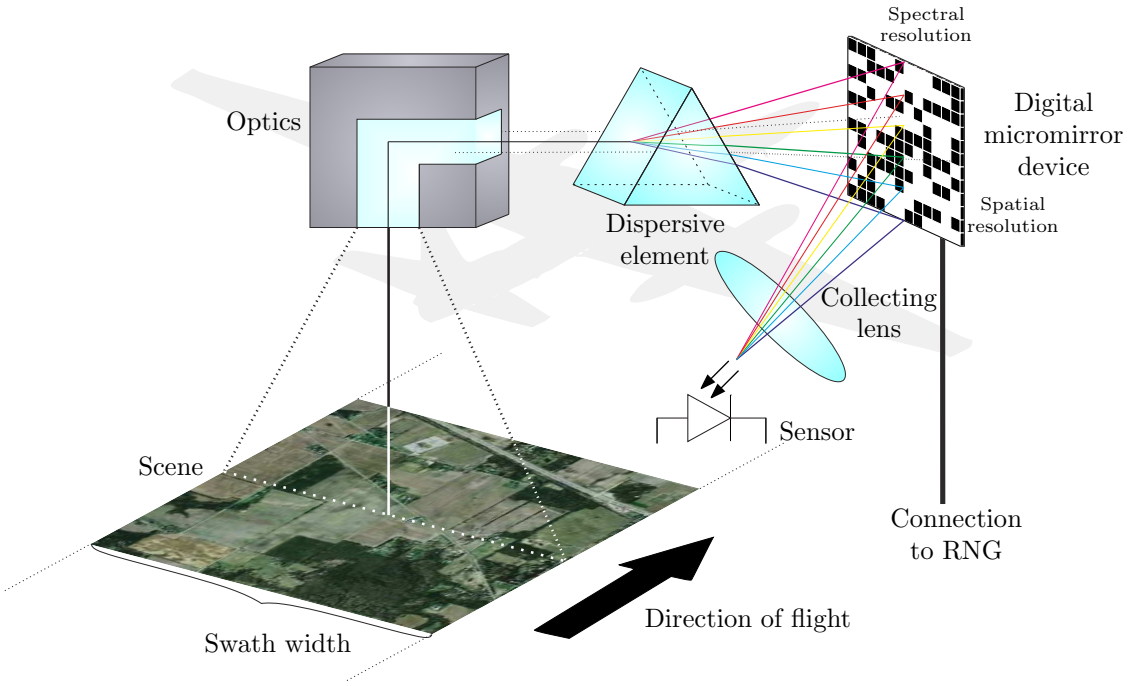


Figure 2: Design of a compressed sensing hyperspectral imaging system using spectral stacking

Since it is a dense matrix, it occupies large amounts of memory and therefore storing N_1 different matrices of that size would be intractable. For this reason, the same measurement matrix is used for all layers. Note that a random number generator (RNG) is used to set the entries of $\mathbf{\Phi}$.

3.1.1. Two-dimensional reconstruction algorithm

For convenience, the spatial-spectral layers of the transformed image, i.e. the transform coefficients $\mathbf{S} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$, are vectorized yielding $\mathbf{S}_{\text{vec}} = (\text{vec}(\mathbf{S}_1), \ldots, \text{vec}(\mathbf{S}_{N_1}))^T \in \mathbb{R}^{N_1 \times N_{23}}$ with $N_{23} = N_2 N_3$. The reconstructed transformed i-th layer $\hat{\mathbf{S}}_i \in \mathbb{R}^{N_2 \times N_3}$ with $1 \leq i \leq N_1$ is obtained by solving the following optimization problem:

$$\begin{aligned} \text{vec}(\hat{\mathbf{S}}_i) = &\arg\min_{\mathbf{S}_i} \|\text{vec}(\mathbf{S}_i)\|_1 \\ &\text{s.t. } \|\mathbf{\Phi}\mathbf{\Psi}^{\mathbf{T}}\text{vec}(\mathbf{S}_i) - \mathbf{y}_i\|_2 \leq \beta_{l_1} \end{aligned} \tag{6}$$

where $\mathbf{y}_i$ denotes the observation vector of the respective layer. The parameter β_{l_1} regularizes between sparsity and exact reconstruction, which is required since the data is not perfectly sparse. The orthonormal transform into a sparse domain, $\mathbf{\Psi}$, is the two-dimensional discrete wavelet transform based on the Haar-wavelet [27]. Throughout this paper, we will only consider the Haar-wavelet since the adaptive direct sampling approach is also based on it for

reasons explained in Section 4.

After the coefficients $\hat{\mathbf{S}}$ have been estimated, the estimate of the original image $\hat{\mathbf{X}}$ is obtained by the inverse transform of each layer, i.e. $\mathrm{vec}(\hat{\mathbf{X}}_i) = \mathbf{\Psi^T}\mathrm{vec}(\hat{\mathbf{S}}_i)$.

3.1.2. Three-dimensional reconstruction algorithm

In order to exploit additional information by considering all layers at once, we may simply change the reconstruction problem. Using the TV norm, the hyperspectral image is then reconstructed by

$$\begin{aligned} \hat{\mathbf{X}} = &\arg\min_{\mathbf{X}} \|\mathbf{X}\|_{\mathrm{TV}} \\ &\text{s.t. } \|\mathbf{\Phi}\mathrm{vec}(\mathbf{X}_i) - \mathbf{y}_i\|_2 \leq \beta_{\mathrm{TV}} \quad \forall\, i \in [1, N_1] \end{aligned} \tag{7}$$

where β_{TV} is the regularization parameter and the TV norm of three-dimensional data is given by

$$\begin{aligned} \|\mathbf{X}\|_{\mathrm{TV}} = &\sum_{i}^{N_1}\sum_{j}^{N_2}\sum_{k}^{N_3}\Big(\left(X_{i+1,j,k} - X_{i,j,k}\right)^2 \\ &+ \left(X_{i,j+1,k} - X_{i,j,k}\right)^2 \\ &+ \left(X_{i,j,k+1} - X_{i,j,k}\right)^2\Big)^{\frac{1}{2}} \\ = &\sum_{i}^{N_1}\sum_{j}^{N_2}\sum_{k}^{N_3}\|(\nabla\mathbf{X})_{i,j,k}\|_2 \end{aligned} \tag{8}$$

with $X_{i,j,k}$ denoting the (i,j,k)-th pixel. Note that the acquisition is still performed on a two-dimensional layer, i.e. the full correlation of the hyperspectral image cube is not exploited during acquisition.

It is known that the total variation is supposed to work well only on piece-wise smooth signals. We are aware of the fact that the spectral curves are usually not piece-wise smooth. However, we suggest the use of the TV-norm for two reasons: first, as will be seen later, the achieved reconstructions are of high quality. Second, the TV-norm is similar to the Haar-wavelet in the sense that both are related to the gradient of the image. The Haar wavelet, on the other hand, is used for Adaptive Direct Sampling (ADS), since it is optimal with respect to the number of samples. Thus, a fair comparison of the methods is provided.

3.2. Spatial stacking

Instead of decomposing the hyperspectral image cube into layers of spectral-spatial images, it is also possible to treat the image as a stack of spatial-spatial images where each band is individually captured (Fig. 1(b)). Mathematically, both approaches are very similar. However, the acquisition of individual spatial-spatial images is far more complex in practice.

Both approaches present, independently of the reconstruction algorithm, an intuitive approach for CS in HSI. Unfortunately, they are suboptimal since only parts of the correlation within the hyperspectral image cube is exploited during the acquisition of the data.

4. Structured Measurement Matrices for HSI

In the previous section, we presented a system based on the single-pixel camera [25, 26] that uses CS to capture hyperspectral images. While in this setup the layers are captured independently, the correlation within the hyperspectral image cube was only exploited by the three-dimensional reconstruction formulation.

In this section, we investigate adaptive direct sampling (ADS) [28] and explain how this method can be adapted to hyperspectral imaging by fully exploiting the correlation within the hyperspectral image. In contrast to CS, ADS implies structured, deterministic measurements that reflect the coefficients of the discrete wavelet transform (DWT) based on the Haar wavelet. In contrast to the single-pixel camera [25, 26], Deutsch *et al.* [28] exploit the characteristics of the DWT for a direct sampling method, referred to as adaptive direct sampling (ADS). It is possible to compute a single coefficient by means of the DMD array that represents a row of the DWT matrix. It is known from compression algorithms such as JPEG2000 that images are highly compressible in the wavelet domain [29]. However, there are two major differences between ADS and compression based on transform coding: in this application, the coefficients of the transform are optically computed and it is not known in advance which are the most important for a good approximation of the image. Thus, finding an efficient sampling scheme that predicts which samples should be captured is crucial for the success of the system.

The advantages of this method are numerous: first, the reconstruction of the image is performed simply by an inverse transform, where no optimization algorithm is required. Thus, reconstruction is faster, easier and less expensive compared to CS. Second, its adaptive nature allows for a monotonic increase of the image quality with every measurement, allowing to obtain a preview of the scenery at any time. In the worst case, after having measured all coefficients of the signal, it can be reconstructed without loss [28]. This guarantee is given by CS only when signals are truly sparse in the given domain. As we are dealing with real world signals, the DWT coefficients are only approximately sparse and therefore, CS will probably capture the image with a certain loss of information.

However, it remains to be asked how to choose the coefficients to be measured. For each measurement, the aim is to map a row of the DWT on the DMD array, similar to the measurement matrix $\mathbf{\Phi}$ in compressive imaging [26]. Though there exist some DMDs that allow for different settings of the deflection angle, most available DMDs are only able to present binary states. In the sequel, we therefore detail how the transform matrix is decomposed into positive binary matrices.

4.1. Characteristics of the discrete wavelet transform

Generally, the DWT coefficients $\mathbf{y}$ of the signal $\mathbf{x}$ with length N_1 are obtained by

$$\mathbf{y} = \mathbf{H}_{N_1}^{(1\mathrm{D})}\mathbf{x} \tag{9}$$

where $\mathbf{H}_{N_1}^{(1\mathrm{D})}$ denotes the discrete wavelet transform matrix which is obtained by writing the $\log_2(N_1)$ filter stages of the high and low pass filters in matrix notation. An algorithm for computing this matrix is explained in [30].
In the sequel, the meaning of the elements of $\mathbf{H}_{N_1}^{(1\mathrm{D})}$ play an important role. For convenience, Eq. (12) exemplarily shows the matrix for a signal length of $N_1 = 8$ where the high and low pass filters are given by

$$\mathbf{g} = \begin{bmatrix} g_0 & g_1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \end{bmatrix} \tag{10}$$

$$\text{and} \quad \mathbf{h} = \begin{bmatrix} h_0 & h_1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix} \tag{11}$$

using the Haar-wavelet.

It can be seen from Eq. (11) and Eq. (12) that the first row contains only coefficients of the low pass filter and therefore, it has a special meaning: it computes the signal scaled down to only one coefficient, i.e. this coefficient represents the image at the coarsest resolution. The coefficients obtained by means of the other rows add detail information. This observation is exploited by the tree structure of the DWT [31] where the transformed signal $\mathbf{y}$ is split into levels of detail. Accordingly, the coarsest value is the most important one since all other values depend on it.

4.2. Decomposing the discrete wavelet transform matrix

Using the Haar-wavelet, the m-th row of $\mathbf{H}^{(1\mathrm{D})}$ contains zeros (sparsity) and a value $\pm d_m^{(1\mathrm{D})} \in \mathbb{R}$. The reason for this property is given by the Haar-wavelet that consists of only ± 1 and the scaling d_0. Considering the absolute value of the elements, the number of different elements reduces to only two. Furthermore, since all elements in a row describe the same number of passed

$$\mathbf{H}_8^{(1\mathrm{D})} = \begin{bmatrix} h_1^3 & h_1^2 h_0 & h_0 h_1^2 & h_0^2 h_1 & h_0 h_1^2 & h_1 h_0^2 & h_0^2 h_1 & h_0^3 \\ g_1 h_1^2 & g_1 h_1 h_0 & g_1 h_0 h_1 & g_1 h_0^2 & g_0 h_1^2 & g_0 h_1 h_0 & g_0 h_0 h_1 & g_0 h_0^2 \\ g_1 h_1 & g_1 h_0 & g_0 h_1 & g_0 h_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & g_1 h_1 & g_1 h_0 & g_0 h_1 & g_0 h_0 \\ g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & g_1 & g_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g_1 & g_0 \end{bmatrix} \tag{12}$$

filter stages and all filters have the same scale ($d_0 = \frac{1}{\sqrt{2}}$), the stated property is proven. Consequently, the absolute value of the scale of each row can be stored in $\mathbf{d}^{(1\mathrm{D})} = (d_1^{(1\mathrm{D})}, \ldots, d_{N_1}^{(1\mathrm{D})})^T$ so that

$$\mathbf{H}^{(1\mathrm{D})} = \mathrm{diag}\Big(\mathbf{d}^{(1\mathrm{D})}\Big) \left(\mathbf{H}_+^{(1\mathrm{D})} - \mathbf{H}_-^{(1\mathrm{D})}\right) \tag{13}$$

where the binary matrices $\mathbf{H}_+^{(1\mathrm{D})}$ and $\mathbf{H}_-^{(1\mathrm{D})}$ have entries of value one where $\mathbf{H}^{(1\mathrm{D})}$ is positive or negative, respectively, and zeros otherwise. This decomposition holds for arbitrary dimensions (see Appendix). Thus, two DMD operations are required to capture a single coefficient of the DWT. Note that both $\mathbf{H}_+^{(1\mathrm{D})}$ and $\mathbf{H}_-^{(1\mathrm{D})}$ contain only positive values and, thus, can easily mapped onto a DMD. The subtraction is performed subsequent to the acquisition of the positive and negative part of the coefficient by a processing unit.
The two-dimensional DWT $\mathbf{H}^{(2\mathrm{D})}$ of a vectorized two-dimensional signal is obtained by $\mathbf{H}^{(2\mathrm{D})} = \mathbf{H}_{N_2}^{(1\mathrm{D})} \otimes \mathbf{H}_{N_1}^{(1\mathrm{D})}$ resulting in a sparse matrix with $\otimes$ denoting Kronecker's product.

As we are interested in exploiting the correlation of three-dimensional data, coefficients of the three-dimensional DWT need to be captured. For this purpose, the transform matrix $\mathbf{H}^{(3\mathrm{D})} = \mathbf{H}_{N_3}^{(1\mathrm{D})} \otimes \mathbf{H}^{(2\mathrm{D})}$ is decomposed as follows (see Appendix):

$$\begin{aligned}\mathbf{H}^{(3\mathrm{D})} = \mathrm{diag}\Big(\mathbf{d}^{(3\mathrm{D})}\Big) &\left(\mathbf{H}_+^{(1\mathrm{D})} - \mathbf{H}_-^{(1\mathrm{D})}\right) \\ &\otimes \left(\mathbf{H}_+^{(2\mathrm{D})} - \mathbf{H}_-^{(2\mathrm{D})}\right)\end{aligned} \tag{14}$$

where $\mathbf{d}^{(3\mathrm{D})} \in \mathbb{R}^N$ with $N = N_1 N_2 N_3$ denotes the scaling vector of the three-dimensional DWT. Analog to $\mathbf{H}_+^{(1\mathrm{D})}$ and $\mathbf{H}_-^{(1\mathrm{D})}$, $\mathbf{H}_+^{(2\mathrm{D})}$ and $\mathbf{H}_-^{(2\mathrm{D})}$ represent the locations of the positive and negative elements, respectively.

In Eq. (14), it is important to realize that each element in a row of $\mathbf{H}^{(3\mathrm{D})}$ has the same scaling coefficient. Furthermore, $\mathbf{H}^{(3\mathrm{D})}$ can be split again into a scaling vector and (positive) binary matrices. Thus, this transform can be split into basically three operations: first a two-dimensional transform, second a one-dimensional transform and finally the scaling of the coefficients to preserve the orthonormality of the transform. This decomposition is used by the system design introduced in the following section.

4.3. System design

As already shown, a compressed imaging camera requires only a single sensor element. Since the spatial information of all pixels has to be captured by this sensor to allow subsequent reconstruction, the DMD array collecting the data must be of the same size as the final image. This also applies to hyperspectral imaging if the concept of the single pixel camera is directly transferred. Considering the fact that, besides the spatial, also the spectral information has to fit on the DMD array, a significant trade-off between spatial and spectral resolution

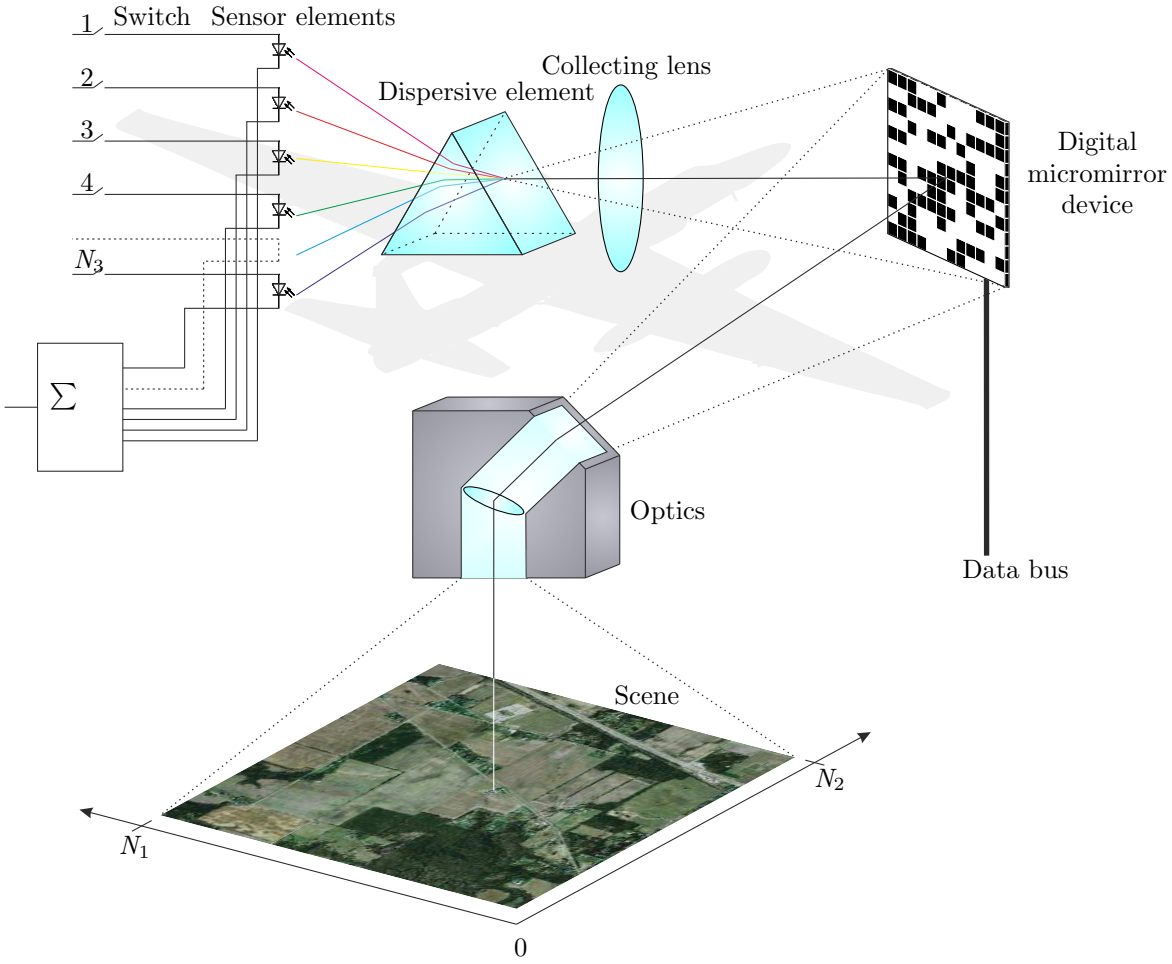


Figure 3: Design of a compressed sensing hyperspectral imaging system using ADS

has to be found if the full correlation within the hyperspectral image is to be exploited for each measurement. Instead, we suggest the following capturing procedure: first, a DMD array is used to encode only the spatial information using a two-dimensional Haar DWT. After the reflected light has been collected by a lens, it is decomposed using a dispersive element, e.g. a prism. The spectral components are then measured by performing a one-dimensional DWT by switching on and off several detectors as depicted in Fig. 3.

4.3.1. Two-dimensional transform

For the previously mentioned reasons, first, the image $\mathbf{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ is only two-dimensionally transformed. As already explained, the two-dimensional DWT can be performed by mapping a row of $\mathbf{H}_+^{(2D)}$ and $\mathbf{H}_-^{(2D)}$ on the DMD array

while the subtraction is performed by a processing unit. For convenience, the spatial layers (bands) are vectorized yielding $\mathbf{X}_{\text{vec}} \in \mathbb{R}^{N_{12} \times N_3}$ with $N_{12} = N_1 N_2$ and thus, the physical operation of the DMD array and lens is described by

$$\mathbf{t}_{m,+}^{(2\text{D})} = \mathbf{h}_{m,+}^{(2\text{D})} \mathbf{X}_{\text{vec}} \tag{15}$$

$$\text{or} \quad \mathbf{t}_{m,-}^{(2\text{D})} = \mathbf{h}_{m,-}^{(2\text{D})} \mathbf{X}_{\text{vec}} \tag{16}$$

with $\mathbf{h}_{m,+}^{(2\text{D})}$ and $\mathbf{h}_{m,-}^{(2\text{D})}$ denoting the m-th row of $\mathbf{H}_{+}^{(2\text{D})}$ or $\mathbf{H}_{-}^{(2\text{D})}$, respectively. In this step, it is important to realize that the $\mathbf{h}_{m,+}^{(2\text{D})}$ and $\mathbf{h}_{m,-}^{(2\text{D})}$ contain only positive binary values as explained in Section 4.2 and thus can easily mapped onto a DMD. In order to obtain the coefficient at row i and column j of $\mathbf{H}^{(2\text{D})}$, m has to be set to $m = i + N_1 (j-1)$. Note that Eq. (15) and Eq. (16) are each performed in a single step for all bands using a single DMD array since the image has not been spectrally decomposed yet, allowing the parallel processing of the bands in incoherent light. In contrast to [20], two DMD operations are required to capture a single coefficient since the sensing matrix itself is not binary. Since the values are subtracted later, the number of stored coefficients is only half the number of measurements of the DMD.

4.3.2. Three-dimensional transform and scaling

After passing the dispersive element, access is gained to the spectral information. Using N_3 spectral sensors, a one-dimensional DWT is performed by summing up and scaling the values measured by the active sensor elements, i.e.

$$\begin{aligned} t_{i,j,k}^{(3\text{D})} = d_n^{(3\text{D})} \Big[&\Big(\Big\langle \mathbf{h}_{k,+}^{(1\text{D})}, \mathbf{t}_{m,+}^{(2\text{D})} \Big\rangle - \Big\langle \mathbf{h}_{k,+}^{(1\text{D})}, \mathbf{t}_{m,-}^{(2\text{D})} \Big\rangle \Big) \\ &- \Big(\Big\langle \mathbf{h}_{k,-}^{(1\text{D})}, \mathbf{t}_{m,+}^{(2\text{D})} \Big\rangle - \Big\langle \mathbf{h}_{k,-}^{(1\text{D})}, \mathbf{t}_{m,-}^{(2\text{D})} \Big\rangle \Big) \Big] \end{aligned} \tag{17}$$

with $\mathbf{h}_{k,+}^{(1\text{D})}$ and $\mathbf{h}_{k,-}^{(1\text{D})}$ denoting the k-th row of $\mathbf{H}_{+}^{(1\text{D})}$ and $\mathbf{H}_{-}^{(1\text{D})}$, respectively. The index n of the element of scaling vector $\mathbf{d}^{(3\text{D})}$ is given by $n = i + N_1 (j-1) + N_1 N_2 (k-1)$. Obviously, for each DMD measurement, two operations of the sensors are required. Since, however, the activation and deactivation of the sensors is much faster than switching the state of the DMDs and the results are immediately summed after the acquisition, this issue will not be considered in the following paragraph. Note that the scaling $d_n^{(3\text{D})}$ is performed by a processing unit.
Using this procedure, the coefficient $t_{i,j,k}^{(3\text{D})}$ of the 3D-DWT transformed image at position (i, j, k) is computed. This is summarized in Fig. 4.

4.4. Sampling strategy

Deutsch *et al.* [28] suggest to maintain a queue Q that contains the indexes of all coefficients that are to be sampled. The key idea is to find those coefficients that exhibit the largest absolute values by means of the tree structure within the DWT. This implies two assumptions: first, the absolute value of a coefficient

Compute $\mathbf{H}_{+}^{(2D)}$, $\mathbf{H}_{-}^{(2D)}$, $\mathbf{H}_{+}^{(1D)}$, $\mathbf{H}_{-}^{(1D)}$ and $\mathbf{d}^{(3D)}$ using Eq. (12) and Eq. (14)

While $z <$ maxSamples

- $z \leftarrow z + 1$
- Determine index (i, j, k) of the coefficient to be sampled according to sampling scheme
- $m \leftarrow i + N_1 (j - 1)$
- $n \leftarrow i + N_1 (j - 1) + N_1 N_2 (k - 1)$
- Set DMD array pattern to m-th row of $\mathbf{H}_{+}^{(2D)}$
  - Set sensor array pattern to n-th row of $\mathbf{H}_{+}^{(1D)}$
  - $\tilde{t} \leftarrow$ output
  - Set sensor array pattern to n-th row of $\mathbf{H}_{-}^{(1D)}$
  - $\tilde{t} \leftarrow \tilde{t}$ - output
- Set DMD array pattern to m-th row of $\mathbf{H}_{-}^{(2D)}$
  - Set sensor array pattern to n-th row of $\mathbf{H}_{+}^{(1D)}$
  - $\tilde{t} \leftarrow \tilde{t}$ - output
  - Set sensor array pattern to n-th row of $\mathbf{H}_{-}^{(1D)}$
  - $\tilde{t} \leftarrow \tilde{t}$ + output
- Store scaled coefficient $t_{i,j,k}^{(3D)} = d_n^{(3D)} \cdot \tilde{t}$

Figure 4: Process of capturing the coefficients by the proposed system

reflects its importance and second, the importance of the coefficients is persistent through the paths of the tree. The algorithm proposed by Deutsch *et al.* [28] works as follows. First, the coarsest value is sampled and the queue is initialized with the children of the coarsest value, i.e.

$$\begin{aligned} Q = \{&(1,1,2)\,,(1,2,1)\,,(1,2,2)\,,(2,1,1)\,,\\ &(2,1,2)\,,(2,2,1)\,,(2,2,2)\}\,. \end{aligned} \tag{18}$$

Then, the position of the next coefficient is fetched from Q which works based on a first-in-first-out principle. The children of a sampled coefficient are added to the queue if either the coefficient is on a coarse level, i.e. below level T_{lev}, or if it is on a higher detail level, and the absolute value of the coefficient exceeds a certain threshold T_{val}. Note that the indexes of the children of a coefficient at the position (i,j,k) can be calculated by [31]

$$\begin{aligned} \text{ch}(i,j,k) = \{&(2i-1,2j-1,2k-1)\,,\\ &(2i,2j-1,2k-1)\,,\\ &(2i-1,2j,2k-1)\,,\\ &(2i,2j,2k-1)\,,\\ &(2i-1,2j-1,2k)\,,\\ &(2i,2j-1,2k)\,,\\ &(2i-1,2j,2k)\,,\\ &(2i,2j,2k)\}\,. \end{aligned} \tag{19}$$

if the children in this image exist, i.e. $1 \leq i \leq \frac{N_1}{2}$, $1 \leq j \leq \frac{N_2}{2}$, and $1 \leq k \leq \frac{N_3}{2}$.

In general, it is unknown which values of the parameters T_{lev} and T_{val} yield the best results. Since the values of the parameters that result in high-quality reconstructions significantly change from image to image, this method may be difficult to apply in practice. Especially T_{val} has a remarkable influence on the captured image.

Therefore, we suggest a novel, alternative sampling scheme overcoming the need to set any parameter. Instead of using fixed preset thresholds, we propose an adaptive scheme that chooses the coefficients whose parents have the largest absolute value. An ordered queue, which can be implemented by means of a red-black tree [32, 33] or a simple binary tree, is used that allows for efficiently storing, finding and deleting the maximum value. Binary trees present an alternative to red-black trees but suffer from the fact that the tree might degrade to a list when an ordered set of samples is inserted. Thus, the complexity for searching the maximum raises to $\mathcal{O}(n)$.

In contrast, a balanced tree, such as the red-black tree, avoids this by balancing the tree under certain conditions. Red-black trees fulfill certain requirements (red-black properties) that guarantee the minimum height of the tree [34]. If these requirements are violated due to inserting or deleting an element, the imbalance is recognized and handled by rotating the elements so that the

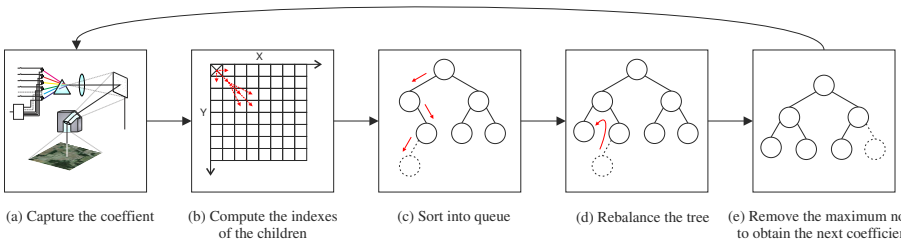


Figure 5: Demonstration of the sampling scheme using the adapted ADS approach

red-black properties are restored. Here, it is important to note that the time required for reorganizing the tree does not depend on the number of elements or the tree height and is therefore of constant complexity. Thanks to the minimum height of the tree, the important operations of insertion, deleting and finding the maximum are achieved in a time complexity of $\mathcal{O}(\log n)$.

Using the ordered queue, we overcome the challenging task to finding a suitable threshold for each image. Furthermore, this procedure is very flexible and does not require sampling the coefficients up to a certain detail level. In this sense, it is a simple, yet completely automatic, self-adapting approach.

The sampling scheme based on a balanced tree is demonstrated in Fig. 5. Following this description, first, (a) a coefficient is captured. By Eq. (19), (b) the children of this sample are determined and (c) sorted into the queue. If necessary, (d) the tree is balanced. The index of the sample found in the right most leaf (e) will be captured next and therefore removed from Q.

4.5. Storage requirements

Using the compressive imaging based methods, we only need to store the captured samples. We assume that the same seed is used for the (pseudo) random number generator (RNG) in the base station as in the mobile system so that the measurement matrices required for the reconstruction can be easily reproduced. Of course, the same RNG has to be used in the mobile device as in the base station.
With ADS, not only the coefficients but also their locations have to be stored. As explained in Section 4.3, the number of stored samples for ADS is only half the number of the captured samples, so that

$$\#\{\text{ADS stored samples}\} = \frac{1}{2}\#\{\text{measured samples}\} \tag{20}$$

where $\#\{\cdot\}$ denotes the number of elements in the respective set.

In the following, we do not consider quantization of the coefficients as this would increase the computational complexity. Storing the position of the samples roughly doubles the required memory space. For a deeper analysis, we have to be more specific about a possible implementation. A simple option is to store the three coordinates of the variable in three variables, of e.g. 8 bits and the

Table 1: Example for coding the position of an ADS coefficient using a single integer variable of 32 bits

Bits	Interpretation	Maximum value
31-20	x-coordinate	4096
19-8	y-coordinate	4096
7-0	z-coordinate (band)	256

measured sample in a 32 bit variable. Using unsigned values, this would allow a maximum image size of $256 \times 256 \times 256$ pixels which is clearly too small for real hyperspectral images. An efficient possibility to handle realistic image sizes is to store the position in a single integer variable of e.g. 32 bits that might be coded as shown in Tab. 1. Of course, the coding can be adapted according to the specifications of the system (i.e. image size and number of bands).

Assuming that a measured sample is stored in a variable of the same size as the index variable, we can conclude that the required storage space by ADS (D_{ADS}) is equal to CS (D_{CS}), i.e. $D_{\mathrm{ADS}} = D_{\mathrm{CS}}$.

The stored data can additionally be compressed by e.g. Huffman or arithmetic coding [35] to save further transmission bandwidth. As the main concern of this paper is to show how hyperspectral images can be efficiently captured and the use of these compression techniques is straight forward, we will not go into details of this intermediate step.

5. Experimental Results

5.1. Data sets

The proposed methods have been evaluated in terms of image quality on numerous datasets. As, ultimately, we are mainly interested in the classification accuracy for HSI, in the following part of this section, we consider the *Indian Pines*, *University of Pavia*, *Salinas* and the *2013 IEEE GRSS Data Fusion Contest* data sets as their ground truth was available to us.

5.1.1. Indian Pines

The *Indian Pines* (IP) image was captured by AVIRIS during a flight a few miles west of Lafayette, Indiana, USA in June, 1992 [36, 37]. The size of the image is 145×145 pixels and 224 bands showing mainly vegetation at a resolution of 20 $\frac{\mathrm{m}}{\mathrm{pixel}}$. Note that noisy bands have not been removed in out experiments. Fig. 6 shows the real color image of the scene and exemplarily the 20^{th} band of the hyperspectral image. The ground truth reveals that basically 16 different materials occur in the scene, where most of them stem from agriculture.

5.1.2. Salinas

The *Salinas* scene (SAL), captured by AVIRIS in October 1998, shows the Salinas Valley, California, USA. The size of the image is 512×217 samples and 224 bands. The image is only as at-sensor radiance data available and noisy

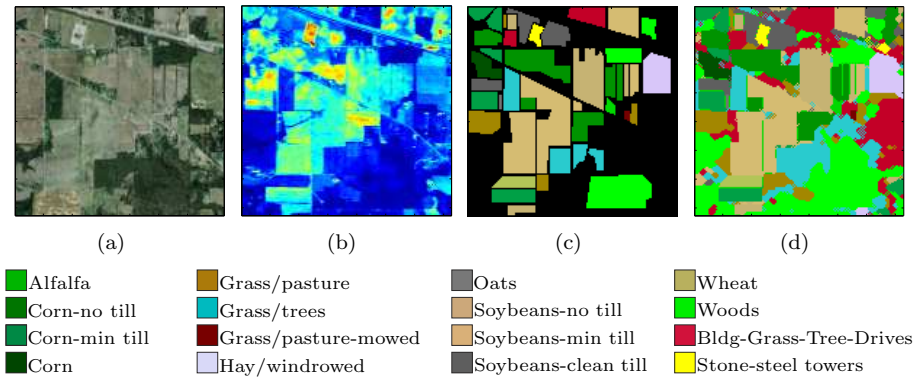


Figure 6: (a) True color image of the *Indian Pines* data set, (b) the 20^{th} band of the hyperspectral image, (c) the ground truth mapping, and (d) final classification result

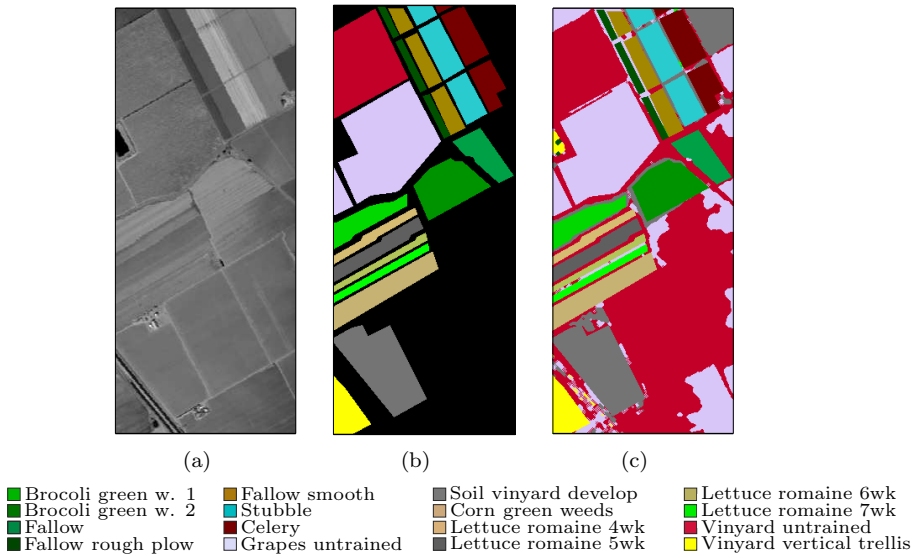
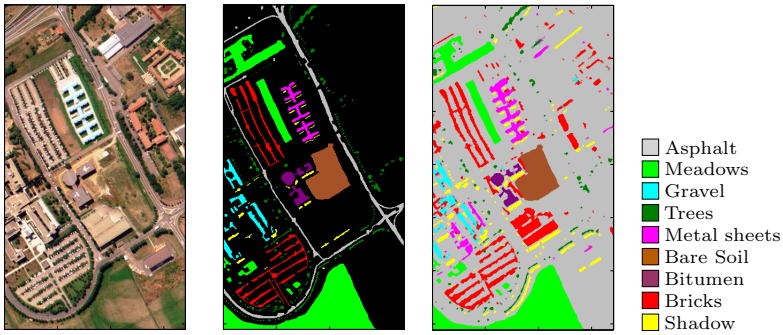


Figure 7: (a) 100^{th} band of the *Salinas* data set, (b) the ground truth mapping and (c) the final classification result



(a) (b) (c)

Figure 8: (a) True color image of the *University of Pavia* data set, (b) the ground truth mapping, and (c) the final classification result

bands have not been removed. Similar to the Indian Pines image, the Salinas image shows mainly different types of vegetation but at a much higher spatial resolution of 3.7 $\frac{\text{m}}{\text{pixel}}$ [38].

5.1.3. University of Pavia

The *University of Pavia* image (UP) was taken by the Digital Airborne Imaging Spectrometer (DAIS) and Reflective Optics System Imaging Spectrometer (ROSIS). The flight was conducted by the "Deutsche Zentrum für Luft- und Raumfahrt" (DLR) in 2002. Both systems produced together an image of 610×340 pixels and 115 bands, where twelve bands are removed due to strong noise, resulting in 103 bands. During the flight, ground and air measurements were conducted to allow atmospheric correction of the image. In contrast to the *Indian Pines* image, the spatial resolution is considerably higher at $1.3\frac{\text{m}}{\text{pixel}}$. Since this image shows an urban area, the nine materials in this scene stem from very different groups, i.e. metal, plant or asphalt.

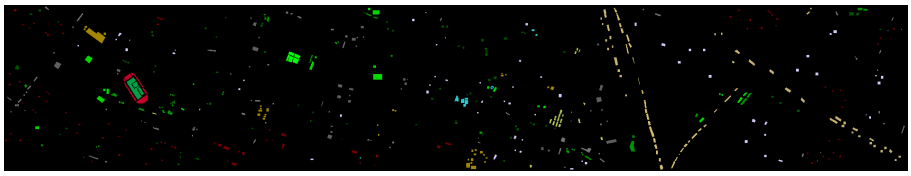
5.1.4. 2013 IEEE GRSS Data Fusion Contest data set

The *2013 IEEE GRSS Data Fusion Contest* (GRSS) data set was captured over the University of Houston campus, Houston, Texas, USA in June 2012. The data was collected by the NSF-funded Center for Airborne Laser Mapping. The image has a size of 349×1905 pixels with 144 bands ranging from 380nm ot 1050nm. The spatial resolution is slightly lower as in the the UP image. A major problem with the classification of this image is a large shadow of a cloud in the right part of the image, causing distorted signatures [39, 40].

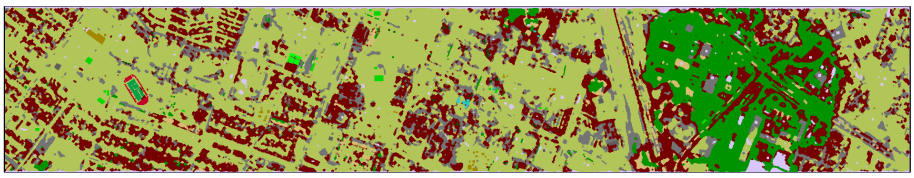
5.2. Sampling and reconstruction

In this section, the quality of the reconstructed images is investigated. Due to computational demands, the images are decomposed into cubic blocks with block size N_b, which are independently reconstructed, allowing for feasible runtimes

(a)



(b)



(c)

Figure 9: (a) True color image of the *IEEE GRSS* data set,d (b) the 100th band of the hyperspectral image, and (c) the ground truth mapping

of the algorithms. For convenience, we simply used overlapping blocks at the borders of the image. The shown results present the averaged outcomes over all blocks. Note that noisy bands in case of *Indian Pines* and *Salinas* were not removed.

As reconstructing complete images may take up to several hours, we first show results how the different methods perform at a fixed block size of $N_\mathrm{b} = 32$ pixels with different subsampling rates R_s. We define the subsampling rate R_s as

$$R_\mathrm{s} = \frac{D_\mathrm{ADS}}{D_\mathrm{ORG}} = \frac{D_\mathrm{CS}}{D_\mathrm{ORG}} \tag{21}$$

where D_ORG denotes the required storage space for the original (raw) image. Note that we compare the storage space, meaning that we would obtain a perfect reconstruction for ADS if $R_\mathrm{s} = 2$, as explained in Section 4.5.

In a different experiment, the impact of the block size with a fixed subsampling rate $R_\mathrm{s} = 0.4$ is investigated. Both tests are used to find promising methods, block sizes and subsampling rates that are further examined in the third experiment where complete hyperspectral images are reconstructed. For evaluation, we choose the *Indian Pines* data set due to its popularity. A second order cone program [41] is used together with a two-dimensional Haar DWT for the minimization of the l_1-norm (L1). Furthermore, the behavior of TV minimization is examined [42]. Note that all algorithms use the same random values for the measurement matrix. The image quality of the reconstructed images are measured by the peak signal-to-noise ratio (PSNR). The PSNR, as it is based on the mean-square error, measures the difference between two signals. Here, it is important to note that all methods used for the reconstruction in this work are based on the idea that the image is sparse in some domain. Since noise is often assumed to be uncorrelated (white), it cannot be sparsely represented. Thus, all algorithms exhibit a denoising effect on the data which means that the captured image is denoised at no extra cost. But for this reason, the PSNR values have to be treated with care since they measure the difference with the original image – which surely exists due to the denoising.

5.2.1. Subsampling test

In Fig. 10(a), the quality of the reconstructed images measured by the PSNR for different subsampling rates are shown. The results reveal that CS works well for HSI. The PSNR of the reconstructed image in all cases monotonically increases when more samples are captured. From this figure, we can generally state that for the *Indian Pines* dataset, a high PSNR is, generally, given at $R_\mathrm{s} = 0.3$ where all methods show a PSNR of more than 30 dB. The best performance is shown by ADS, where even a subsampling rate of $R_\mathrm{s} = 0.1$ is sufficient to obtain an image with a PSNR of 34.53 dB. The worst performing method is the reconstruction based on the l_1-norm which does not achieve a PSNR higher than 40 dB for any tested subsampling rate. Further, even if $R_\mathrm{s} = 1.0$, the image is not perfectly reconstructed in any case.

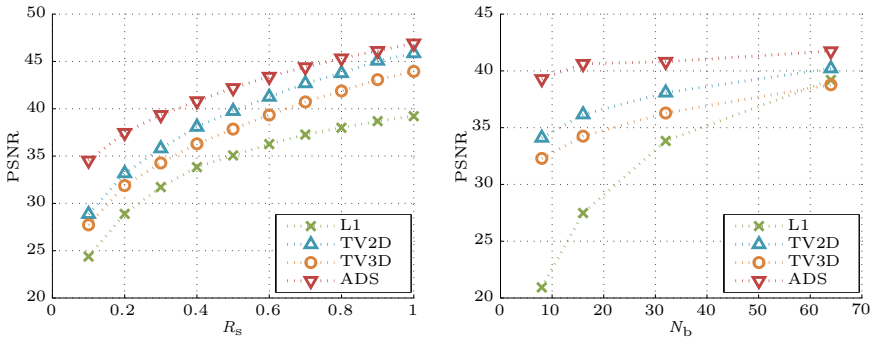


Figure 10: Results of the PSNR at different (a) subsampling rates and (b) block sizes

5.2.2. Block size test

The second test provides information about the change of the image quality for different block sizes. The subsampling rate is constant at 0.4 and the size of the blocks is varied from 8, 16, 32 to 64 pixels per edge of the each block (sizes with a basis of two are chosen because of the discrete wavelet transforms). Due to computational limitations, reconstructing blocks of sizes above 64 pixels is intractable for most cases.

Fig. 10(b) shows the PSNR results of the reconstructed images for different block sizes. In general, an increase of the block size leads to a higher PSNR. While ADS benefits only slightly from increasing the size from 8 to 64 pixels (+2.4 dB), the effect is remarkable for reconstructing by means of the l_1-norm (+18.2 dB).

5.2.3. Reconstructing Complete Images

From the previous results, we can expect high-quality reconstructions at $R_s = 0.4$. Since we are interested in reducing R_s, results for $R_s = 0.1$ are also presented. The block size is fixed to $N_b = 32$ pixels.

In order to provide an impression about the results presented in Tab. 2, the reconstruction of the 20^{th} band of the *Indian Pines* image is depicted in Fig. 11. Using a subsampling rate of 0.4, all methods result in images of a high quality which are difficult to distinguish from each other.

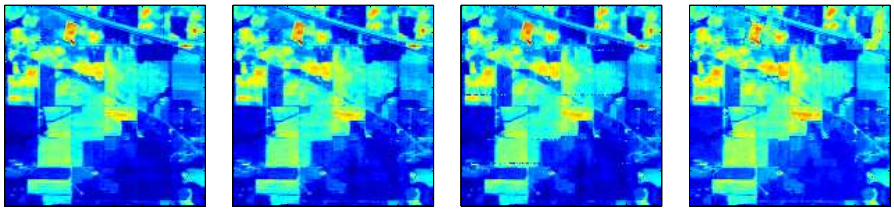
However, differences become more apparent when the subsampling rate is as low as 0.1. In case of ADS, even with $R_s = 0.1$, there are only few artifacts due to the Haar wavelet, whereas L1 and TV-2D clearly suffer from the layer-wise reconstruction. For TV-3D, smoother results are obtained.

5.3. Classification

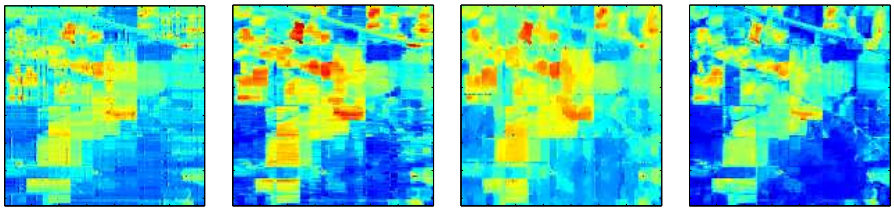
The quality of a hyperspectral image is only one factor to assess the quality of the reconstruction. In fact, hyperspectral images are often directly processed by computers to extract valuable information and are thus not viewed by humans.

Table 2: Results of the complete reconstruction of the Indian Pines and University of Pavia image

(R=0.4/R=0.1)	Indian Pines PSNR [dB]	Salinas PSNR [dB]	University of Pavia PSNR [dB]	GRSS PSNR [dB]
L1	33.83 / 24.38	29.44 / 20.4	33.49 / 23.67	**36.71** / 22.79
TV-2D	38.08 / 28.86	35.48 / 24.14	34.73 / 24.46	33.47 / 23.92
TV-3D	36.29 / 27.73	34.37 / 24.31	30.03 / 22.81	30.38 / 22.9
ADS	**40.81** / **34.53**	**41.07** / **34.49**	**35.37** / **27.56**	35.71 / **27.72**



(a) L1, $R_s = 0.4$ (b) TV-2D, $R_s = 0.4$ (c) TV-3D, $R_s = 0.4$ (d) ADS, $R_s = 0.4$



(e) L1, $R_s = 0.1$ (f) TV-2D, $R_s = 0.1$ (g) TV-3D, $R_s = 0.1$ (h) ADS, $R_s = 0.1$

Figure 11: Reconstruction of the *Indian Pines* image (20^{th} band)

Consequently, classification assessment plays an important role as it provides further information about the usefulness of the presented methods.
In this section, we answer the question how the classification accuracy is affected by the reconstruction. Instead of enhancing existing methods, a pixelwise classification by support vector machines (SVM) [43] is used to allow for a better comparison to existing methods, e.g. [44, 45, 46]. Besides, SVMs have successfully been used in many other areas, e.g. sonar or through-the-wall imaging [47, 48].

First, cross-validation is performed to find and evaluate the parameters that are optimal with respect to the false classification rate. For this purpose, the ground truth data is split into two sets. The first and larger set is used to perform model selection using cross-validation [49]. The second part is then used to evaluate the obtained parameters and to estimate the misclassification error. By this, the data sets for the test and for model selection are strictly independent and therefore provide a more realistic estimate of the error. In the following, two metrics are used to judge the quality of the classification: (i) the overall accuracy (OA), i.e. the ratio of correctly classified data and the total number of data and (ii) the average per class accuracy (CA). Furthermore, all methods and images are investigated for subsampling rates of 0.1 and 0.4. Note that the results of conventional sensing, i.e. the original image, is denoted by ORG.
Finally, all parts of the images are classified, i.e. also those parts where no ground truth is available, to give an impression of the performance of the presented methods.

5.3.1. Results of cross-validation

As depicted in Fig. 12, all methods exhibit similar accuracies at $R_\mathrm{s} = 0.4$. Except for *Indian Pines*, ADS performs best for all data sets, close to ORG. The differences between the performances of the methods become more significant when the subsampling rate is reduced to only 0.1. For all data sets, the accuracies obtained by classifying images reconstructed by L1 remarkably drop off. Especially for the *GRSS* and *University of Pavia* data sets, the compressive sensing methods clearly suffer strongly from the reduced subsampling rate, while only ADS is nearly unaffected. The results form the *Salinas* dataset should be treated with care since there is only little structure in the image, meaning that the blocks can be highly sparsely represented. Thus, there is only little change in the classification accuracy when reducing R_s from 0.4 to 0.1 for this data set.

5.3.2. Complete image classification

Finally, complete images are classified by the methods yielding the highest overall accuracies in order to provide an impression about their performance. For this purpose, all samples where class labels are available are used to conduct a model selection by means of cross-validation. The optimal parameters and chosen subsampling rates found for each image are given in Tab. 3. The results of the classification are depicted in Fig. 6(d) and Fig. 7(c)-Fig. 9(c). Note that SVMs are used to estimate class probabilities that are further processed

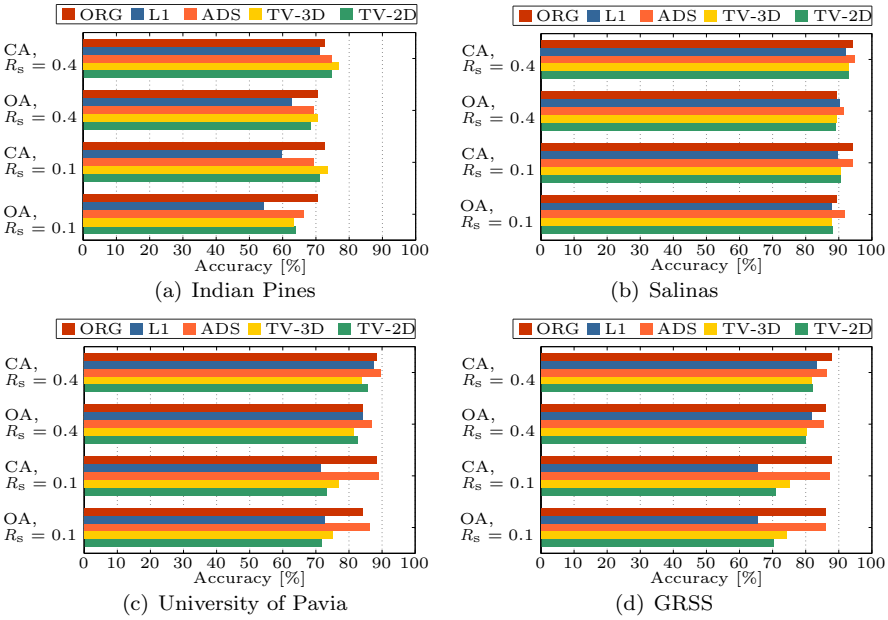


Figure 12: Cross-validation results of the *Indian Pines*, *Salinas*, *University of Pavia* and *GRSS*, block size: 32 pixels

by a simplified iterative scheme similar to [45] so as to exploit the spatial information within the image for each pixel. Comparing the classified image with the original images shows clearly that the presented methods, especially ADS, yield acceptable classification results.

6. Discussion

As shown in the previous section, the spectral stacking approach using either TV-2D or TV-3D for reconstruction as well as the ADS system present capable designs that are able to provide results comparable to conventional acquisition

Table 3: Optimal parameters found by means of a cross-validation for the datasets and best capturing methods

Dataset	Indian Pines	Salinas	University of Pavia	GRSS
Capturing Method	TV-3D	ADS	ADS	ADS
R_s	0.4	0.1	0.4	0.1
CV-OA	84.4%	93.59%	96.5%	90.3%
Parameters of the SVM	$C = 128$, $\gamma = 0.125$	$C = 256$, $\gamma = 10^{-3}$	$C = 128$, $\gamma = 1.0$	$C = 1024$, $\gamma = 0.125$

systems. This result has been verified by comparing image quality and classification accuracy. Reconstructing images by L1 is clearly suboptimal, especially for small block sizes, as indicated by low PSNR values and classification accuracies. As for conventional images, the TV methods represent a far more effective approach.

Thus, two different systems have been presented. But which one is preferable? In terms of classification, this question can be clearly answered as ADS performs better than TV in most cases, especially for low subsampling rates. However, if the main purpose of the system is to maximize the quality of the image, the stacking approach may provide slightly more accurate results. Interestingly, TV often showed a better performance than its three-dimensional counterpart, though the latter is expected to exploit more information for the reconstruction. A reason for this might consist in the ability of TV-3D to provide solutions that are sparser than those of TV with respect to the gradient. Therefore, small details may not be accordingly reconstructed. The fact that TV is designed for piece-wise smooth curves might come into play here as well. Another benefit of ADS consists in its fast reconstruction, consisting of only an inverse transform, i.e. a matrix multiplication. In contrast, CS is based on an optimization problem that is usually iteratively solved. Thus, in our experiments, ADS required only a few seconds while most CS algorithms took several minutes up to a few hours, depending on the algorithm and image size.

For computational reasons, the images were split into blocks of 32 pixels per edge in the experimental part. However, optimizing the implementation of the algorithms and using appropriate hardware, larger block sizes up to an edge length equal to the number of bands would be achievable. As it is shown in the size tests, increasing the block size while keeping the subsampling rate constant results in a higher image quality. Thus, we can assume that increasing the block size also increases sparsity, i.e. fewer samples are required.

Wavelet-based compression algorithms use more advanced wavelets than the Haar-wavelet to improve the compression rate [29, 27]. In contrast, the Haar-wavelet exhibits properties that make it relatively easy to use for the acquisition of hyperspectral images. First, this wavelet consists only of two elements, possessing the same absolute value which leads to the second property that it can be easily written by positive binary matrices and a scaling vector. The latter property enables Haar-wavelets to be represented by a binary DMD array.

7. Conclusions and Outlook

Two systems were presented allowing direct capture of compressed hyperspectral images. First, the setup for spectral stacking with different reconstruction problems was explained. In addition, adaptive direct sampling based on the idea that images are highly compressible in the wavelet domain was adapted to hyperspectral imaging. For this purpose, another system design was presented as well as a new automatic, self-adapting sampling scheme. In the experimental part, the achieved image qualities and classification accuracies demonstrate that both systems are able to produce images that are very similar to those

conventionally captured at a rate of 0.4. However, reducing the rate to only 0.1, adaptive direct sampling (ADS) outperformed the other methods in all cases. Surprisingly, we found that reducing the compression rate using ADS improved the classification accuracy in few cases. This motivates to look for a different sampling strategy that is optimized with respect to classification instead of image quality.

The question arises, whether other types of wavelets can also be used in the presented system. First of all, they lead to a different matrix of the discrete wavelet transform, i.e. it can no longer be decomposed into two positive binary matrices. Hence, one option would be to decompose the new transform matrix until only binary elements occur where the scaling is separately saved. Consequently, the number of measurements increases with each decomposition and it is questionable whether this increase can be balanced by the saved samples. This approach remains interesting if the system is to be optimized to reduce only the number of stored coefficients. Another option consists in arrays of gray-value digital micromirror devices (DMD) that are capable of representing values up to 2^{10}. These devices are relatively new and are therefore not very common. However, a resolution of 1024 different values might be insufficient to represent the values of other wavelets. If the approximated values are too imprecise, the property of orthonormality will no longer hold, leading to a biased reconstruction. However, finding an alternative to DMDs, i.e. using a different light modulator, might simplify this problem.

Considering the fact that all methods decrease the computational load on the mobile system, is there any reason not to use CS based methods or ADS? A severe problem is that those methods have a high exposure time. Similar to the single-pixel camera, ADS, as well as the other CS based methods, samples the image in the time domain. Imagine a moderate image size of $256 \times 256 \times 220$ where 1,441,792 samples need to be captured using a subsampling rate of 0.1. Current standard DMD arrays are able to switch their states about 5000 times in a second. Let the system use a high-end DMD that is at least four times faster, capturing the image takes roughly 72 seconds in case of ADS. This is a significant disadvantage especially for spaceborne systems that orbit the earth at high speeds. But even for airborne imagery this poses a problem. With the *pushbroom* scanning method in conventional remote hyperspectral imaging, aircrafts are used as moving platforms capturing a single line of the image at a time. As this method prevents full exploitation of the correlation within the data, it cannot be used for ADS. Of course, a trivial solution could consist in waiting for the development of even faster DMD arrays. This will take many years and increase costs of the system. Another option is to optimize the sampling scheme so that even less samples are required. It is known that hyperspectral images are highly compressible using the wavelet transform, and other works showed that often a subsampling rate of only 0.01 leads to very good results. If the sampling scheme was able to find these coefficients, the image in this example would be captured in only 7.2 seconds.
Furthermore, the scene of interest could be split into small blocks as we did in the evaluation of reconstructing complete images. Each block would be of the

size of all required bands, the swath width and an arbitrary length that can be adapted to the speed of the platform. Even at this early stage of development, the presented method is useful in practical systems. A possible application consists in mounting this design to helicopters or gyroplanes that allow for longer exposure times. This is especially interesting for autonomous or remotely controlled systems similar to drones or unmanned aerial vehicles for the purpose of local operations such as recon or rescue missions. Obviously, the problem of the high exposure time is not as significant for stationary systems that become more and more interesting for industrial purposes.

Finally, we conclude based on the classification accuracies, computation time as well as required memory capacity that the system based on ADS is generally preferable. However, if the system is designed to maximize image quality, spatial stacking using TV reconstruction remains an interesting approach.

Appendix A.

The proof of Eq. (14) is as follows: the matrix $\mathbf{H}^{(2D)}$ of the two-dimensional DWT is obtained by

$$\mathbf{H}^{(2D)} = \mathbf{H}^{(1D)} \otimes \mathbf{H}^{(1D)}. \tag{A.1}$$

As explained in Section 4.2, $\mathbf{H}^{(1D)}$ can be decomposed so that

$$\begin{aligned} \mathbf{H}^{(2D)} = &\left[\mathrm{diag}\left(\mathbf{d}^{(1D)}\right)\ \left(\mathbf{H}_{+}^{(1D)} - \mathbf{H}_{-}^{(1D)}\right)\right] \\ &\otimes \left[\mathrm{diag}\left(\mathbf{d}^{(1D)}\right)\ \left(\mathbf{H}_{+}^{(1D)} - \mathbf{H}_{-}^{(1D)}\right)\right] \end{aligned} \tag{A.2}$$

where $\mathbf{d}^{(1D)}$ is the scaling vector of the one-dimensional DWT. The positive binary matrices $\mathbf{H}_{+}^{(1D)}$ and $\mathbf{H}_{-}^{(1D)}$ indicate positive and negative entries of the respective transform matrices. Using the mixed-product property of Kronecker's product [50] leads to

$$\begin{aligned} \mathbf{H}^{(2D)} = &\left[\mathrm{diag}\left(\mathbf{d}^{(1D)}\right) \otimes \mathrm{diag}\left(\mathbf{d}^{(1D)}\right)\right] \\ &\left[\left(\mathbf{H}_{+}^{(1D)} - \mathbf{H}_{-}^{(1D)}\right) \otimes \left(\mathbf{H}_{+}^{(1D)} - \mathbf{H}_{-}^{(1D)}\right)\right]. \end{aligned} \tag{A.3}$$

Since $\mathrm{diag}\left(\mathbf{d}^{(1D)}\right) \otimes \mathrm{diag}\left(\mathbf{d}^{(1D)}\right)$ results again in a diagonal matrix, Eq. (A.3) can be written as

$$\begin{aligned} \mathbf{H}^{(2D)} = &\mathrm{diag}\left(\mathbf{d}^{(2D)}\right)\ \left[\left(\mathbf{H}_{+}^{(1D)} - \mathbf{H}_{-}^{(1D)}\right)\right. \\ &\left. \otimes\ \left(\mathbf{H}_{+}^{(1D)} - \mathbf{H}_{-}^{(1D)}\right)\right] \end{aligned} \tag{A.4}$$

with $\mathbf{d}^{(2D)} = \mathbf{d}^{(1D)} \otimes \mathbf{d}^{(1D)}$. Decomposing Kronecker's product of the binary matrices into two positive binary matrices $\mathbf{H}_{+}^{(2D)}$ and $\mathbf{H}_{-}^{(2D)}$ representing the

positive and negative values, gives finally

$$\mathbf{H}^{(2\mathrm{D})} = \mathrm{diag}\Big(\mathbf{d}^{(2\mathrm{D})}\Big)\ \Big[\Big(\mathbf{H}_{+}^{(2\mathrm{D})} - \mathbf{H}_{-}^{(2\mathrm{D})}\Big)\Big]. \tag{A.5}$$

The decomposition of the three-dimensional DWT matrix $\mathbf{H}^{(3\mathrm{D})} = \mathbf{H}^{(1\mathrm{D})} \otimes \mathbf{H}^{(2\mathrm{D})}$ is computed likewise. Thus, this procedure can be repeated to show that any L-dimensional DWT matrix with $L > 1$ can be written as $\mathbf{H}^{(L\mathrm{D})} = \mathrm{diag}\Big(\mathbf{d}^{(L\mathrm{D})}\Big)\ \Big[\Big(\mathbf{H}_{+}^{(L\mathrm{D})} - \mathbf{H}_{-}^{(L\mathrm{D})}\Big)\Big]$.

Acknowledgment

The authors would like to thank Prof. D. Landgrebe from Purdue University, U.S., for kindly providing the Indiana's *Indian Pines* data set, Prof. P. Gamba from University of Pavia, Italy, for providing the *University of Pavia* image [51], and Dr. L. Johnson and Dr. J. A. Gualtieri for providing the AVIRIS *Salinas* data set. Further, the authors would like to thank the Hyperspectral Image Analysis group and the NSF Funded Center for Airborne Laser Mapping (NCALM) at the University of Houston for providing the *2013 IEEE GRSS Data Fusion Contest* data set used in this study, and the IEEE GRSS Data Fusion Technical Committee for organizing the 2013 Data Fusion Contest [40].

References

[1] G. A. Shaw, H. K. Burke, Spectral imaging for remote sensing, Lincoln Laboratory Journal 14 (1) (2003) 3–28.

[2] R. A. Schowengerdt, Remote Sensing: Models and Methods for Image Processing, 3rd Edition, Elsevier, 2007.

[3] F. M. Lacar, M. M. Lewis, I. T. Grierson, Use of hyperspectral imagery for mapping grape varieties in the Barossa Valley, South Australia, in: Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS 2001), Vol. 6, 2001, pp. 2875–2877.

[4] D. Haboudane, Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture, Remote Sensing of Environment 90 (3) (2004) 337–352.

[5] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. Winter, A. Schaum, A. Stocker, Anomaly detection from hyperspectral imagery, IEEE Signal Processing Magazine 19 (1) (2002) 58–69.

[6] S. Matteoli, M. Diani, G. Corsini, A tutorial overview of anomaly detection in hyperspectral images, IEEE Aerospace and Electronic Systems Magazine 25 (7) (2010) 5–28.

[7] A. Zare, J. Bolton, P. Gader, M. Schatten, Vegetation mapping for landmine detection using long-wave hyperspectral imagery, IEEE Transactions on Geoscience and Remote Sensing 46 (1) (2008) 172–178.

[8] A. Thomas, J. Cathcart, Applications of grid pattern matching to the detection of buried landmines, IEEE Transactions on Geoscience and Remote Sensing 48 (9) (2010) 3465–3470.

[9] A. Gowen, C. O'Donnell, P. Cullen, G. Downey, J. Frias, Hyperspectral imaging - an emerging process analytical tool for food quality and safety control, Trends in Food Science & Technology 18 (12) (2007) 590–598.

[10] A. Picon, O. Ghita, P. Iriondo, A. Bereciartua, P. Whelan, Automation of waste recycling using hyperspectral image analysis, in: IEEE Conference on Emerging Technologies and Factory Automation (ETFA), 2010, 2010, pp. 1–4.

[11] E. J. Candès, J. K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Communications on Pure and Applied Mathematics 59 (8) (2006) 1207–1223.

[12] D. Donoho, Compressed sensing, IEEE Transactions on Information Theory 52 (4) (2006) 1289–1306.

[13] J. Trzasko, A. Manduca, Highly undersampled magnetic resonance image reconstruction via homotopic l_0-minimization, IEEE Transactions on Medical Imaging 28 (1) (2009) 106 –121.

[14] M. Tello, P. López-Dekker, J. Mallorquí and, A novel strategy for radar imaging based on compressive sensing, IEEE Transactions on Geoscience and Remote Sensing 48 (12) (2010) 4285 –4295.

[15] M. Leigsnering, C. Debes, A. M. Zoubir, Compressive sensing in through-the-wall radar imaging, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011), 2011, pp. 4008–4011.

[16] M. Golbabaee, S. Arberet, P. Vandergheynst, Multichannel compressed sensing via source separation for hyperspectral images, in: Proceedings of the European Signal Processing Conference (EUSIPCO 2010), 2010.

[17] C. Li, T. Sun, K. Kelly, Y. Zhang, A compressive sensing and unmixing scheme for hyperspectral data processing, IEEE Transactions on Image Processing 21 (3) (2012) 1200–1210.

[18] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, T. J. Schulz, Single-shot compressive spectral imaging with a dual-disperser architecture, Optics Express 15 (21) (2007) 14013–14027.

[19] A. Wagadarikar, R. John, R. Willett, D. Brady, Single disperser design for coded aperture snapshot spectral imaging, Applied Optics 47 (10) (2008) B44–B51.

[20] V. Studer, J. Bobin, M. Chahid, H. S. Mousavi, E. Candes, M. Dahan, Compressive fluorescence microscopy for biological and hyperspectral imaging, in: Proc. National Academy of Sciences, Vol. 109, 2012, pp. E1679–E1687.

[21] A. Jerri, The Shannon sampling theorem: Its various extensions and applications: A tutorial review, Proceedings of the IEEE 65 (11) (1977) 1565–1596.

[22] E. J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Transactions on Information Theory 52 (2) (2006) 489–509.

[23] R. G. Baraniuk, Compressive sensing [lecture notes], IEEE Signal Processing Magazine 24 (4) (2007) 118–121.

[24] E. J. Candès, J. K. Romberg, L1-Magic: Recovery of sparse signals, http://www.acm.caltech.edu/l1magic/.

[25] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, R. Baraniuk, An architecture for compressive imaging, in: IEEE International Conference on Image Processing, 2006, pp. 1273–1276.

[26] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, R. G. Baraniuk, A new compressive imaging camera architecture using optical-domain compression, in: Proc. of Computational Imaging IV at SPIE Electronic Imaging, 2006, pp. 43–52.

[27] S. Mallat, A Wavelet Tour of Signal Processing, Second Edition (Wavelet Analysis & Its Applications), 2nd Edition, Academic Press, 1999.

[28] S. Deutsch, A. Averbuch, S. Dekel, Adaptive compressed image sensing based on wavelet modeling and direct sampling, in: International conference on Sampling Theory and Applications, Vol. 8, 2009.

[29] D. S. Taubman, M. W. Marcellin, Jpeg2000: Image compression fundamentals, standards and practice, Journal of Electronic Imaging 11 (2) (2002) 286–287.

[30] B. Vidakovic, Statistical Modeling by Wavelets, Wiley, New York, 1999.

[31] J. M. Shapiro, An embedded hierarchical image coder using zerotrees of wavelet coefficients, in: Proc. DCC '93. Data Compression Conference, 1993, pp. 214–223.

[32] R. Bayer, Symmetric binary B-trees: Data structure and maintenance algorithms, Acta Informatica 1 (1972) 290–306.

[33] L. J. Guibas, R. Sedgewick, A dichromatic framework for balanced trees, in: 19th Annual Symposium on Foundations of Computer Science, 1978, pp. 8–21.

[34] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, Introduction to Algorithms, 2nd Edition, The MIT Press, 2001.

[35] R. C. Gonzalez, R. E. Woods, Digital Image Processing (3rd Edition), 3rd Edition, Prentice Hall, 2007.

[36] AVIRIS NW Indiana's Indian Pines 1992 data set [Online], available at *ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C.lan* (original files) and *ftp://ftp.ecn.purdue.edu/biehl/PCMultiSpec/ThyFiles.zip* (ground truth).

[37] S. Tadjudin, D. A. Landgrebe, Covariance estimation with limited training samples, IEEE Transactions on Geoscience and Remote Sensing 37 (4) (1999) 2113–2118.

[38] A. Gualtieri, S. R. Chettri, R. Cromp, L. Johnson, Support Vector Machine classifiers as applied to AVIRIS data (1999).

[39] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pizurica, S. Gautama, W. Philips, S. Prasad, Q. Du, F. Pacifici, Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS Data Fusion Contest, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

[40] 2013 IEEE GRSS Data Fusion Contest [Online], *http://www.grss-ieee.org/community/technical-committees/data-fusion/*.

[41] J. Sturm, Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones, Optimization Methods and Software 11–12 (1999) 625–653, version 1.05 available from `http://fewcal.kub.nl/sturm`.

[42] C. Li, An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing, Master's thesis, RICE University (2009).

[43] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (2011) 27:1–27:27, software available at *http://www.csie.ntu.edu.tw/∼cjlin/libsvm*.

[44] F. Melgani, L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, IEEE Transactions on Geoscience and Remote Sensing 42 (8) (2004) 1778–1790.

[45] Y. Tarabalka, M. Fauvel, J. Chanussot, J. A. Benediktsson, SVM- and MRF-based method for accurate classification of hyperspectral images, IEEE Geoscience and Remote Sensing Letters 7 (4) (2010) 736–740.

[46] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, J. C. Tilton, Multiple spectral–spatial classification approach for hyperspectral data, IEEE Transactions on Geoscience and Remote Sensing 48 (11) (2010) 4122–4132.

[47] R. Fandos, C. Debes, A. M. Zoubir, Resampling methods for quality assessment of classifier performance and optimal number of features, Signal Processing 93 (11) (2013) 2956–2968.

[48] C. Debes, J. Hahn, A. Zoubir, M. Amin, Target discrimination and classification in through-the-wall radar imaging, IEEE Transactions on Signal Processing 59 (10) (2011) 4664–4676.

[49] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Inc., 2006.

[50] K. B. Petersen, M. S. Pedersen, The matrix cookbook, version 20081110 (2008).

[51] P. Gamba, A collection of data for urban area characterization, in: Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS 2004), Vol. 1, 2004.