On the impact of Tensor Completion in the Classification of Undersampled Hyperspectral Imagery

Michalis Giannopoulos*†, Grigorios Tsagkatakis†, and Panagiotis Tsakalides*†
*Department of Computer Science, University of Crete
†Institute of Computer Science, Foundation for Research and Technology Hellas (FORTH)
Heraklion, 70013, Greece

***Abstract*—Typical HSI sensors employ scanning along certain dimensions in order to acquire the hyperspectral data cube. Snapshot Spectral Imaging architectures associate a particular spectral band with each pixel, achieving high temporal sampling rates at a lower spatial resolution. In this work, we study the problem of efficient estimation of missing hyperspectral measurements and we evaluate the impact of the reconstruction quality on the subsequent task of classification. We explore two cutting edge techniques for undersampled signal recovery, namely matrix and tensor completion, and we evaluate their performance on hyperspectral data recovery. Furthermore, we quantify the effects of the reconstruction error on state-of-the-art machine learning algorithms via metrics such as classification accuracy and F1-score. The results demonstrate that robust and efficient classification is feasible, even from a substantially reduced number of measurements being available, especially when emerging deep learning approaches are adopted. Moreover, significant gains are obtained when exploring higher order structural information via tensor modelling, as compared to low order matrix-based methods.**

I. Introduction

Hyperspectral Imaging (HSI) has demonstrated great potential in remote sensing, having an established presence in Earth Observation [1], while gaining momentum in terrestrial applications as well [2]. HSI involves the acquisition of complex information across the electromagnetic spectrum, capturing the spectral content of each pixel in the scene and facilitating operations like detecting objects, characterizing materials, and identifying processes.

Each hyperspectral image is naturally represented as a three-dimensional structure, called a hypercube, where two dimensions are spatial and one is spectral. In many cases, a significant number of elements are missing from hypercubes either due to errors or by design. One particular example of missing measurements is the case of Snapshot Spectral Imaging (SSI). SSI architectures rely on the use of spectrally resolved detector arrays where each pixel is associated with a specific spectral band, allowing the acquisition of a full hyperspectral cube from a single exposure [3]. Unfortunately, to achieve high temporal resolution imaging, SSI architectures must sacrifice spatial resolution since only a small subset of pixels observe each spectral band. It comes as no surprise that the performance of the ensuing classification process depends on the efficiency of the missing measurements recovery mechanism that is employed.

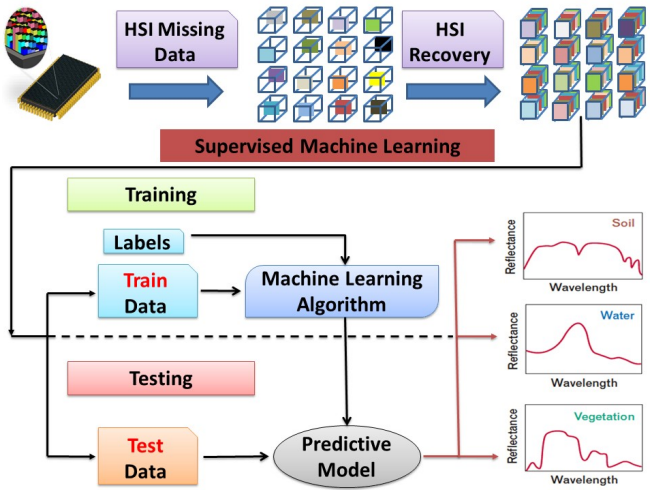

Fig. 1. The proposed framework. Matrix-based and Tensor-based reconstruction and subsequent classification of hyperspectral data.

In this work, we investigate the impact of HSI recovery from a small number of SSI observations on scene classification. We introduce a complete framework, illustrated in Fig. 1, for data structuring, reconstruction, and classification of the overall HSI processing chain in the presence of missing values. We consider two distinct approaches for structuring the data; as two-dimensional matrices and as three-dimensional tensors. Our goal is to quantify the quality of recovery and classification of the full set of observations, through the application of two state-of-the-art recovery approaches, namely Matrix Completion (MC) and Tensor Completion (TC). The key contributions of this paper can be summarized as follows:

- Quantify the potential of missing HSI measurements' reconstruction by adopting MC and TC approaches
- Assess whether further structural information of the data can lead to better and more accurate reconstruction
- Quantify the impact of reconstruction on HSI data classification

While the problems of recovery via MC [4] and classification of SSI data [5] have been explored before, the

ISBN 978-90-827970-1-5

novelty of this work lays in a comprehensive evaluation of the impact of completing *higher-order* structures on classification performance.

II. Recovery and classification with missing measurements

Bearing in mind that the end goal of HSI systems usually refers to machine learning tasks (e.g. supervised classification), the existence of missing measurements in the (train and test) dataset could easily put in jeopardy the whole framework of classification, necessitating the introduction of efficient data imputation techniques.

Given the high-dimensional nature of the hyperspectral data-cube, it seems reasonable to treat it as a third-order tensor structure. Tensors constitute generalizations of vectors and matrices that encode high dimensional structural information [6]. Quite common is the strategy of flattening high-dimensional arrays into two-way matrix structures, in order to perform matrix-based processing. In this case, the dimensions of the matrix correspond to pixels (the two spatial dimensions stacked in a "long" vector) and to spectral bands.

Let us consider a HSI data matrix $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$, where n_1 is the number of pixels and n_2 is the number of spectral bands. Matrix completion tries to recover all its entries from a partially observed fraction of them. More formally, let Ω be the set of known indices (i_1, i_2) corresponding to the available measurements, whereas the linear map $\mathcal{A}$ is defined as an operator setting the values of all unknown indices to zero

$$\mathcal{A}(\mathbf{M}) = \begin{cases} m_{i_1 i_2}, & \text{if } (i_1, i_2) \in \Omega \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

In [7] it was shown that recovery of the missing values from a low rank matrix $\mathbf{M}$ is possible by solving the following rank minimization problem:

$$\begin{aligned} &\underset{\mathbf{X}}{\text{minimize}} \ rank(\mathbf{X}) \\ &\text{subject to} \ \ \mathcal{A}(\mathbf{X}) = \mathcal{A}(\mathbf{M}). \end{aligned} \tag{2}$$

Although rank minimization can recover the matrix, it is impractical for real-life problems due to its NP-hard nature. Fortunately, it has been shown that the nuclear norm, *i.e*, the sum of the singular values, can serve as a proxy to the rank. Thus, the optimization problem in (2) can be reformulated according to

$$\begin{aligned} &\underset{\mathbf{X}}{\text{minimize}} \ \ \|\mathbf{X}\|_* \\ &\text{subject to} \ \ \mathcal{A}(\mathbf{X}) = \mathcal{A}(\mathbf{M}). \end{aligned} \tag{3}$$

To better take advantage of the inherent correlation structure in a HSI data-cube, we herein propose to address the problem of missing measurements by employing a tensor completion approach. In the HSI case, the aforementioned SSI data collection process leading to missing measurements results in an under-sampled $[n_1] \times [n_2] \times [n_3]$ tensor $\boldsymbol{\mathcal{T}}$ (where in this case n_1 and n_2 stand for the spatial dimensions of the data-cube and n_3 for its spectral one), which we wish to fully populate using a fraction k of its measured entries.

Equation (3) can be extended to handle higher-order tensors, by solving the following optimization problem in order to estimate the lowest-rank tensor $\boldsymbol{\mathcal{X}}$ which agrees with the available data

$$\begin{aligned} &\underset{\boldsymbol{\mathcal{X}}}{\text{minimize}} \ \ \|\boldsymbol{\mathcal{X}}\|_* \\ &\text{subject to} \ \ \mathcal{A}(\boldsymbol{\mathcal{X}}) = \mathcal{A}(\boldsymbol{\mathcal{T}}) \end{aligned} \tag{4}$$

where Ω is the index set (i_1, i_2, i_3) of observed entries, and the linear map $\mathcal{A}$ is defined, as before, as a random projection operator keeping the entries in Ω and zeroing out others; that is

$$\mathcal{A}(\boldsymbol{\mathcal{T}}) = \begin{cases} \tau_{i_1 i_2 i_3}, & \text{if } (i_1, i_2, i_3) \in \Omega \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

However, the optimization regime is now tougher than before, as the tensor nuclear norm is not defined as the tightest convex relaxation of the tensor rank, as was the case with matrices. Adopting the approach proposed in [8], one can define the tensor nuclear norm as follows:

$$\|\boldsymbol{\mathcal{X}}\|_* = \sum_{i=1}^{n} \alpha_i \|\mathbf{X}_{(i)}\|_* \tag{6}$$

where n depicts the order-mode of the tensor (i.e. $n = 3$ in our HSI case), and α_i's are weights satisfying $\alpha_i \geq 0$ and $\sum_{i=1}^{n} \alpha_i = 1$. Thus, the nuclear norm for a general tensor case can be defined as the convex combination of the nuclear norms of all matrices unfolded along each of its modes. Under this definition, Equation (4) can be written as

$$\begin{aligned} &\underset{\boldsymbol{\mathcal{X}}}{\text{minimize}} \ \ \sum_{i=1}^{n} \alpha_i \|\mathbf{X}_{(i)}\|_* \\ &\text{subject to} \ \ \mathcal{A}(\boldsymbol{\mathcal{X}}) = \mathcal{A}(\boldsymbol{\mathcal{T}}) \end{aligned} \tag{7}$$

For solving the optimization problem described in (3), we used the Augmented Lagrange Multipliers method (MCALM) [9] as well as the k-Nearest Neighbor Imputation method (KNNI) with $k = 1$ [10]. For tackling the optimization problem in (7), we employed the Low-Rank Tensor Completion using Parallel Matrix Factorization method (TCTMAC) [11].

Once the missing values of the hyperspectral data are recovered, machine learning algorithms such as supervised classification can be applied. In our set-up, we assume that missing measurements occur during both training and testing, since we want to consider the case where the same SSI sensor is employed during both stages. Within the framework of this study, we utilize the data obtained by imputation to train and evaluate the performance of four classes of classifiers. Specifically, we consider the SVM family with linear (LSVM), quadratic (QSVM) and Gaussian (GSVM) kernels; a k-Nearest Neighbours (KNN) classifier with $k = 10$ and Euclidean distance; a Decision Tree (DT) with Gini's diversity index as a split criterion; and finally a Deep Convolutional Neural Network (DCNN) with two-dimensional trainable filters for hierarchical feature learning.

ISBN 978-90-827970-1-5 © EURASIP 2018

III. Experimental Evaluation

In this Section, we report on the experimental evaluation of the proposed scheme using the publicly available Indian Pines [1] annotated hyperspectral dataset. The HSI datacube is composed of 145×145 pixels, over 200 different spectral reflectance bands (in the wavelength range of $400 - 2500$ nanometers), as 24 water absorption bands obtained from the sensor instrument were discarded. The Indian Pines scene contains two-thirds agriculture, and one-third forest or other natural perennial vegetation. Each pixel is labelled using one of sixteen different classes, among which one can find crops, dual lane highways, a rail line, some low density housing, as well as other building structures and smaller roads. It has to be mentioned that supervised training was conducted using the ground truth image of the aforementioned dataset and that background pixels were not considered for classification purposes.

Concerning the employed classifiers, special mention has to be made to the CNNs, a promising type of deep models which produce hierarchically complex features by applying trainable filters and pooling operations on raw input. Training CNNs though is a quite computational costly process, as a convolution of each input with each trainable filter has to take place. To overcome this, prior to the training and the prediction stages of the DCNN classifier, we perform a dimensionality reduction technique across the spectral dimension of the hyperspectral datacube. Based on prior art on this topic [12], we utilized the Randomized PCA method and preserved only the most "informative" 5% of the spectral information (namely, $c = 10$ out of the $C = 200$ spectral bands). As long as the spectral information is condensed, the spatial ones are compensated by forming patches at the vicinity of each pixel in order to take into consideration not only the pixel itself but its closest neighbours as well. The size of each patch, p, is set to 5×5, as bigger sizes do not improve classification accuracy, while computation time increases [12].

The CNN classifier is fed with c inputs of size $p \times p$, and consists of 2 convolutional layers with trainable 3×3 filters. As long as we do not take into consideration any scaling or translation factors, we do not use any max-pooling layers and the number of filters of each layer are set as $C_1 = 3 \times c$ and $C_2 = 3 \times C_1$, respectively [12]. The rest of the employed classifiers are trained with the raw HSI data (i.e. all the available spectral bands).

To evaluate the performance of the end-to-end system, two different quality metrics were considered. The reconstruction quality is evaluated using the Normalized Mean Square Error (NMSE) metric, defined as the mean square error between the fully-populated and the reconstructed data, normalized with respect to the l_2 norm. Regarding the classification process, standard performance metrics such as classification accuracy and the F_1-score are used. Note that to quantify the performance of each step, randomly subsampled observations were first recovered and subsequently used for training and testing. Although in some HSI imagers down-sampling may be performed in a "structured" manner due to the set-up of the acquisition system, in order to be aligned with MC theory in terms of being able to recover the desired solution [7], the sampling set Ω was chosen uniformly at random.

[1] http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

A. Effect of the Training Set Size on the Classification Performance

The objective of this experiment is to assess the performance of the classifiers relative to the size of the training set. A random selection of up to 7174 training examples was considered, and 3075 where utilized for testing. Fig. 2 illustrates the performance of each classifier measured as a function of the number of the training examples.

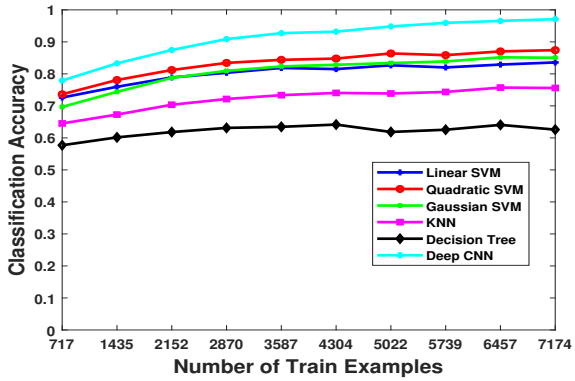

Fig. 2. Classification accuracy w.r.t. the number of the training examples. The more training examples are used, the better the classification accuracy.

The results demonstrate that increasing the number of training examples has a positive effect on the generalization capacity of each classifier, as dictated by the theoretical underpinnings. The majority of the considered classifiers present a stable performance when trained with at least 5000 training examples. It should be noted that the Deep CNN classifier achieves the best performance among the ones employed and it does so with 95% of the spectral information cast away compared to the competitor algorithms, as explained earlier.

B. Effect of the Fill Ratio on the Reconstruction Error

In this set of experiments, we investigate the reconstruction quality of the three considered imputation methods as a function of the number of observed samples. In the matrix-based approaches, namely KNNI and MCALM, the dataset is structured as a 21025×200 matrix, while in the tensor-based modelling a $145 \times 145 \times 200$ tensor was considered. To quantify the performance of the different methods, the fill-ratio metric, f, is defined as the number of the observed samples over the total number of samples in the measurements.

Fig. 3 presents the reconstruction quality in terms of NMSE as a function of f. The results clearly demonstrate that adopting a tensor-based approach offers concrete advantages in terms of reconstruction quality, in contrast to the matrix-based methods. More specifically, the worse performance

ISBN 978-90-827970-1-5 © EURASIP 2018

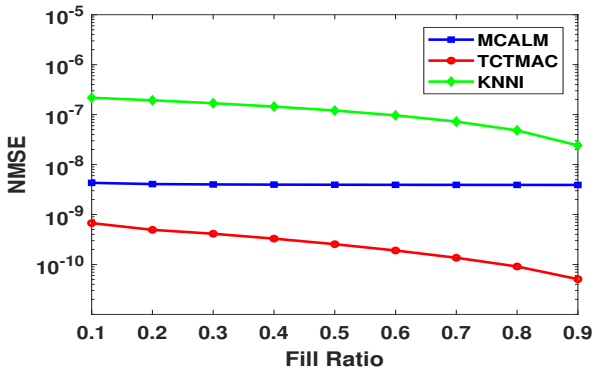

Fig. 3. NMSE w.r.t f for all applied reconstruction methods. Adoption of the higher-order tensor approach clearly leads to more accurate reconstruction.

corresponds to the KNNI method followed by MCALM, while the best performance is obtained by TCTMAC. Even when presented with the smallest number of observations ($f = 0.1$), the performance of TCTMAC is better compared to MCALM at almost full sampling ($f = 0.9$).

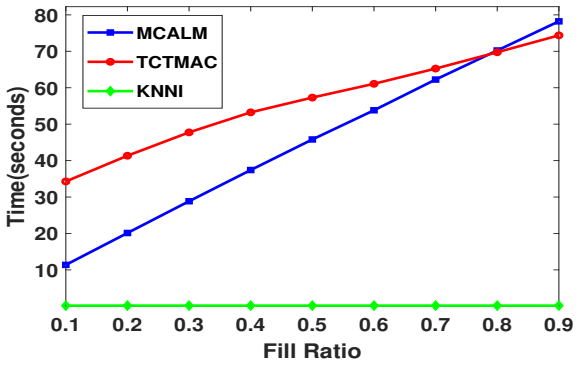

Fig. 4. Computation times with respect to fill ratio f for all reconstruction methods. Tensor-based recovery is the most computational demanding, with a closing gap between TCTMAC and MCALM at larger fill ratios.

Naturally, the improved classification performance of the tensor-based methods does not come for free, as demonstrated by the execution times plots in Fig. 4. However, one should highlight the fact that the gap between TCTMAC and MCALM reconstruction time narrows as the fill-ratio increases, indicating that adopting a higher-order processing structure offers significant merits. Overall, the results clearly demonstrate the merits and reconstruction potential of higher-order modelling.

C. Effects of Reconstruction on Classification Performance

In the final set of experimental results, we incorporate the concept of missing measurements structuring and reconstruction into the overall classification process and we quantify the effect of the reconstruction quality to the resulting classification performance. Fig. 5 illustrates the classification accuracy of all five employed classifiers, for each imputation method, as a function of the fill ratio f.

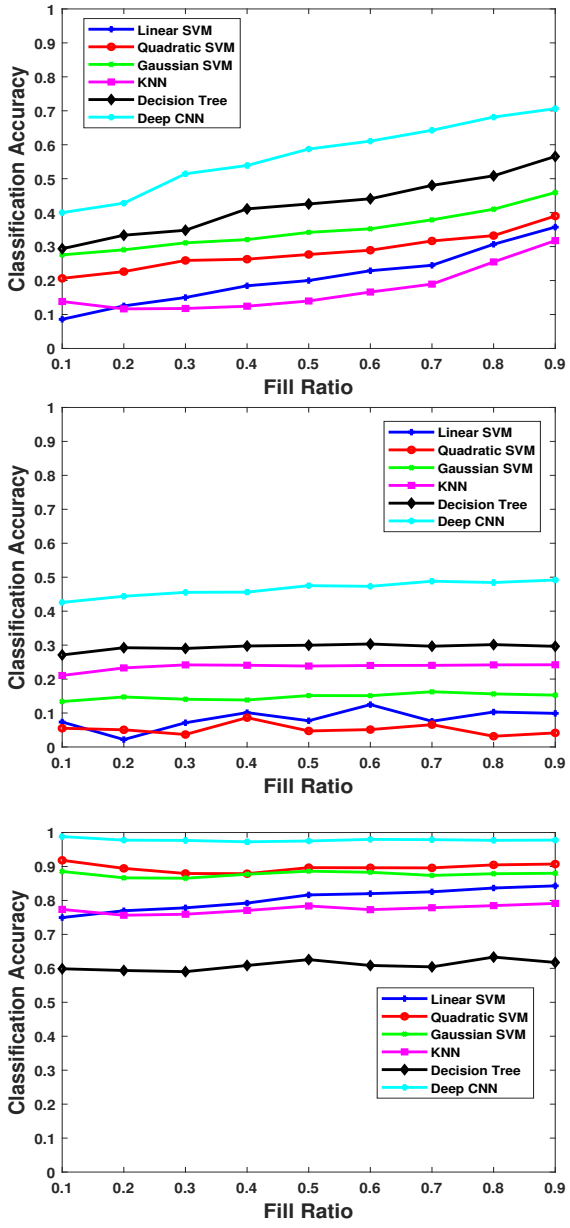

Fig. 5. Classification accuracy for KNNI (top), MCALM (middle) and TCTMAC (bottom) recovery w.r.t. to f. Tensor completed data lead to clearly higher levels of classification accuracy for all employed machine learning algorithms and all different values of f.

An important observation regarding the overall performance of the proposed framework is that, for all employed classifiers, efficient classification accuracy can be achieved by severely undersampled data. This result is in alignment with the results shown in Fig. 3, where we observe that increasing the fill ratio leads to better performance. While for matrix-based methods the performance improves gradually, TCTMAC exhibits remarkable stability and performance, achieving almost optimal classification accuracy even with extremely undersampled data.

ISBN 978-90-827970-1-5 © EURASIP 2018

In order to further support these claims, Table I presents the obtained classification accuracy by all employed classifiers, for each one of the imputation methods and for two representative values of missing data. We focus on the cases of $f = 0.04$ and $f = 0.0625$ sampling due to their significance in actual SSI architectures. The results presented in Table I clearly demonstrate that even when 4% of the observations are available, the tensor-based approach leads to a classification accuracy which is on par to the accuracy achieved using the fully-populated data. On the other hand, the classification accuracy of the matrix-based approaches improves as more data are revealed to the solvers. Nevertheless, both matrix-based methods are clearly outperformed by the tensor-based approach.

TABLE I
CLASSIFICATION PERFORMANCE FOR TWO REPRESENTATIVE SSI FILL RATIOS

f	Method	LSVM	QSVM	GSVM	K-NN	DT	DCNN
0.04	KNNI	0.08	0.15	0.24	0.13	0.27	0.37
	MCALM	0.04	0.01	0.11	0.19	0.25	0.42
	TCTMAC	**0.72**	**0.88**	**0.85**	**0.73**	**0.55**	**0.97**
0.0625	KNNI	0.09	0.19	0.26	0.14	0.28	0.39
	MCALM	0.10	0.06	0.12	0.20	0.25	0.42
	TCTMAC	**0.74**	**0.89**	**0.85**	**0.76**	**0.58**	**0.97**
1	-	0.82	0.91	0.86	0.76	0.62	0.98

Finally, Fig. 6 outlines the performance of the proposed tensor-based imputation approach in terms of the widely used F_1-score machine learning performance metric, for the scenario where 96% of the data are missing.

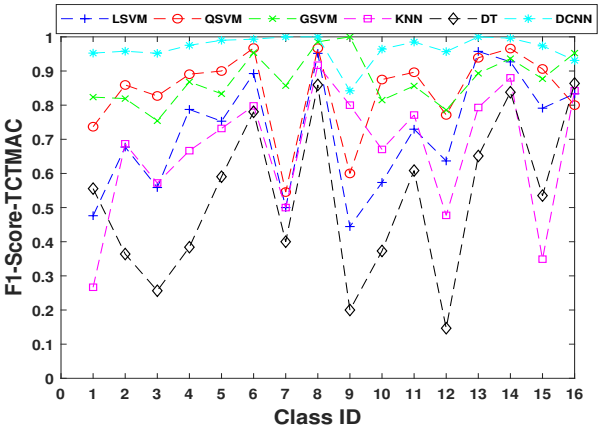

Fig. 6. F_1-score for TCTMAC recovery with $f = 0.04$. Tensor-based imputation has a considerable positive effect, even under quite challenging experimental regimes.

The F_1-score can be interpreted as a weighted average (harmonic mean) of the well-known Precision and Recall metrics. Fig. 6 further confirms our observations regarding the effectiveness of tensor-based recovery in conjunction with supervised classification. Furthermore, it provides a good indication about the efficiency of the proposed scheme in extremely hostile experimental regimes where a significant amount of measurements is missing.

IV. CONCLUSIONS

In this work, we have investigated the effects of missing measurements reconstruction on the classification performance using hyperspectral data. We have focused on two different approaches of imputing missing data: matrix and tensor completion. Based on our experimental findings, we can conclude that excellent classification accuracy is feasible even in the presence of only 4% data observations using tensor-based completion and state-of-the-art DCNNs. Clearly, adopting a tensor-based approach proved to be a better strategy for efficient recovery as compared to the matrix-based imputation algorithms, while its computational complexity, although larger, it is not prohibitive. The results demonstrate that higher-order structuring of the data leads to a significantly better recovery, which in turn translates to an improved classification performance.

V. ACKNOWLEDGMENTS

This work was funded by the General Secretariat for Research and Technology (GSRT), the Hellenic Foundation for Research and Innovation (HFRI), and the DEDALE project, contract no. 665044, within the H2020 Framework Program of the European Commission.

REFERENCES

[1] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 45–54, 2013.

[2] L. M. Dale, A. Thewis, C. Boudry, I. Rotar, P. Dardenne, V. Baeten, and J. A. F. Pierna, "Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: a review," *Applied Spectroscopy Reviews*, vol. 48, no. 2, pp. 142–159, 2013.

[3] B. Geelen, N. Tack, and A. Lambrechts, "A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic," in *Spie Moems-Mems*. International Society for Optics and Photonics, 2014, pp. 89 740L–89 740L.

[4] G. Tsagkatakis, M. Jayapala, B. Geelen, and P. Tsakalides, "Non-negative matrix completion for the enhancement of snapshot mosaic multispectral imagery," in *Electronic Imaging*. SPIE, 2016.

[5] K. fotiadou, G. Tsagkatakis, and P. Tsakalides, "Deep convolutional neural networks for the classification of snapshot mosaic hyperspectral imagery," in *International Symposium on Electronic Imaging*, 2017.

[6] E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos, "Tensors for data mining and data fusion: Models, applications, and scalable algorithms," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 2, p. 16, 2016.

[7] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.

[8] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, 2013.

[9] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[10] E. Acuna and C. Rodriguez, "The treatment of missing values and its effect on classifier accuracy," in *Classification, clustering, and data mining applications*. Springer, 2004, pp. 639–647.

[11] Y. Xu, R. Hao, W. Yin, and Z. Su, "Parallel matrix factorization for low-rank tensor completion," *arXiv preprint arXiv:1312.1254*, 2013.

[12] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*. IEEE, 2015, pp. 4959–4962.

ISBN 978-90-827970-1-5 © EURASIP 2018