AN ANALYSIS OF COLLABORATIVE REPRESENTATION SCHEMES FOR THE CLASSIFICATION OF HYPERSPECTRAL IMAGES

M. Dalla Mura[1], *J. M. Bioucas-Dias*[2], *J. Chanussot*[1,3]

[1]GIPSA-lab, Department Image Signal (DIS),Grenoble Institute of Technology, France
[2]Telecommunications Institute, Instituto Superior Técnico, Lisbon, Portugal
[3]Faculty of Electrical and Computer Engineering, University of Iceland, Iceland

ABSTRACT

Collaborative-based representation classifiers have widely spread in the latest years achieving remarkable results in signal and image processing tasks. In this paper, we consider these approaches for the hyperspectral image classification. Specifically, we focus on collaborative and sparse representation classifiers and we perform an investigation on the role of the different regularizations and constraints that can be considered with respect to the classification performance. In addition, we propose to consider the Nearest Subspace Classifier with regularization which, from the experiments, has proven to be a competitive classification technique. Experimental results have been conducted considering both spectral and spatial features of a real hyperspectral image.

Index Terms— Sparse representation classification, collaborative classification, nearest subspace classifier, hyperspectral imaging, remote sensing

1. INTRODUCTION

Collaborative-based representation classifiers perform a representation of an unlabeled sample as a linear combination of other samples of known label (i.e., linear regression) [1]. The set of labeled samples used as regressors is called *dictionary* and its elements *atoms*. The representation can be modeled in a variational formulation where one seeks an approximation as close as possible to an input signal by finding the weights defining the combination of the atoms. Those weights can be referred to as *code*. Such scheme is called *collaborative* since atoms concur in the representation of a signal. The idea that make this strategy appealing for classification relies on the consideration that a signal is typically similar (read correlated) to samples of its same class [2]. This means that samples of the same class should achieve a more accurate reconstruction of the signal with respect to considering atoms of the other classes. Classification can then be performed, once a sample is reconstructed (i.e., the code is found), by assigning the sample to the class whose atoms that are involved more in the representation (i.e., if considered alone they produce the lowest reconstruction error). In general in the process of identification of the code regularization and constraints can be employed. Regularization based on the ℓ_1 norm of the code has been successfully applied as a way to enforce sparsity on the solution (i.e., favoring few atoms to be active in the representation – in general, obtaining a selection of atoms that are relevant in the classification). Classifiers based on sparse representations (named Sparse Representation Classifiers, SRC) have been proven their effectiveness in several applicative domains of signal and image processing [1]. Focusing on remote sensing, with a particular regard to the analysis of hyperspectral images, sparse techniques have been successfully employed in spectral unmixing [3] and classification [4, 5]. With regard to the latter task, it has been seen that sparsity can cope well with classification with small sample size scenarios [6] such as hyperspectral image classification. SRC has shown to deal very well when processing spatial features for image classification achieving outstanding performances [5]. Although sparse regularization positively influences classification, it has a significant computational complexity. The problem of finding a solution of the reconstruction problem under sparsity needs to be solved numerically and it remains practically tractable for small dictionaries. For this reason, strategies based on other forms of regularization such as those based on ℓ_2 have been proposed in order to make the problem more computationally tractable, especially when closed form solutions exist. Those approaches have been referred to as Collaborative Representation Classifiers (CRCs). Arguments have brought about the greater influence of the collaborative scheme on the performances in classification rather than the type of regularization [7]. Thus, legitimating regularization strategies lighter in terms of computational burden even if less effective in the task of classification with respect to SRC. Collaborative classifiers based on ℓ_2 regularization have been started to appear and show their effectiveness for the classification of hyperspectral images [8, 9].

This paper mainly stems from the work [5] in which SRC was successfully applied in hyperspectral image classification using a set of spatial features. However, little insight was given in that work on the role of the regularization parameters

This work was supported in part by the INS2I JCJC 2015 “IDES” project.

and constraints considered. With this work we want to make such analysis and, in addition, compare the SRC with the CRC strategy in this context. We consider also in this work a Nearest Subspace Classifier (NSC) with regularization as an additional way to restrain the solution space. NSC follows an analogous strategy to S/CRC, with a difference in the representation of the unknown samples. In NSC, the representation is done for each class separately considering only the atoms in the dictionary belonging to that class. Thus, NSC can be thought as a class-wise collaborative representation classifier. NSC has been mainly considered without regularization [6], leading in general to lower performances with respect to regularized schemes. In this work we explore the capabilities of this collaborative strategy but considering regularization.

The remainder of the paper is as follows. A basic introduction of collaborative-based representation classifiers and NSC is given in the following section. Section 3 reports the results of the experimental analysis carried out on a hyperspectral image. Concluding remarks are presented in the last section.

2. COLLABORATIVE-BASED REPRESENTATION CLASSIFIERS

Let us consider an unlabeled sample $\mathbf{x} \in \mathbb{R}^l$ (i.e., in this case, a pixel in a hyperspectral image of l bands) that will be processed by a collaborative-representation classifier. In this scenario, we have at our disposal a dictionary $\mathbf{A}$ composed of n labeled samples (i.e., training set), grouped in m classes: $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_m\}$ with a subdictionary $\mathbf{A}_i$ being composed of the n_i training samples belonging to the $i-$class ($\sum_{i=1}^{m} n_i = n$). The sample $\mathbf{x}$ is then represented as a linear combination of the element of the dictionary in which the weights are given by the vector $\boldsymbol{\alpha}$. Once $\boldsymbol{\alpha}$ is defined, it is possible to estimate the thematic class of $\mathbf{x}$ according to:

$$\widehat{\text{class}}(\mathbf{x}) = \arg \min_{i \in \{1,\ldots,m\}} \|\mathbf{x} - \mathbf{A}_i\boldsymbol{\alpha}_i\|_2. \tag{1}$$

In words, the sample $\mathbf{x}$ is assigned to the class whose elements in the dictionary lead to the least residuals in the reconstruction. The way the vector of coefficients $\boldsymbol{\alpha}$ is derived determines the type of classifier.

2.1. Sparse and collaborative representation

A collaborative representation classifier solves the following problem:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\mathbf{x} - \mathbf{A}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_p, \tag{2}$$

being λ a scalar value modulating the penalty and $\|\boldsymbol{\alpha}\|_p$ the p-norm on the code implementing the regularization term.

For Sparse Representation Classification, $p = 1$ whereas for Collaborative Representation $p = 2$. We underline that here the vector of coefficients in the representation is derived considering all atoms of all classes simultaneously.

2.2. Nearest Subspace Classifier

In NSC a test sample is assigned to the subspace to whom the sample is closest to. So this leads to solve the following problem for every class i:

$$\min_{\boldsymbol{\alpha}_i} \|\mathbf{x} - \mathbf{A}_i\boldsymbol{\alpha}_i\|_2^2. \tag{3}$$

Then the coefficient vector can be constructed as:

$$\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m\}.$$

In this strategy, the coefficients in $\boldsymbol{\alpha}$ are accounted separately for each class. NSC can be seen as a way to automatically avoid the simultaneous contribution of multiple atoms in the dictionary as it might happen for CRC with feeble or without regularization, especially in overdetermined scenarios (i.e., $n > l$). In a way, NSC can induce a form of sparsity since it natively disables some atoms when performing the reconstruction of a signal. However, without regularization NSC might lead to poor representations affecting the classification.

In this work we propose to consider a version of the NSC with regularization. Namely,

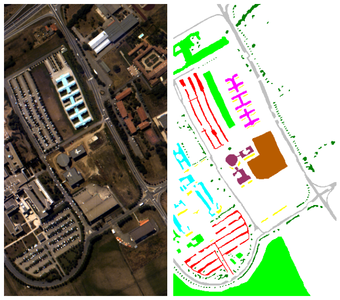
$$\min_{\boldsymbol{\alpha}_i} \|\mathbf{x} - \mathbf{A}_i\boldsymbol{\alpha}_i\|_2^2 + \lambda\|\boldsymbol{\alpha}_i\|_p, \tag{4}$$

As for S/CRC, the regularization based on ℓ_1- or ℓ_2-norms on the code can be coupled with other constraints. For example, non-negativity (NN) can be imposed on the code (i.e., $\boldsymbol{\alpha} \geq 0$). In hyperspectral unmixing, the positivity on the code is related to a physical property of the solution i.e., the non negativity of the abundances [3]. Whereas, in classification such constraint is not associated to a physical meaning but can help in constraining the space of solutions as shown in [5].

3. EXPERIMENTAL RESULTS

3.1. Experimental set-up

The tests are run on a well-known hyperspectral image of the University of Pavia, Italy. The image was collected by the ROSIS optical sensor over the urban area of the University of Pavia, Italy. The flight was operated by the Deutschen Zentrum for Luftund Raumfahrt (DLR, the German Aerospace Agency) in the framework of the HySens project, managed and sponsored by the European Union. The image size in pixels is 610×340, with very high spatial resolution of 1.3 meters per pixel. The number of data channels in the acquired image is 103 (with spectral range from 0.43 to 0.86 μm). Nine thematic land-cover classes were identified in the university campus: Trees, Asphalt, Bitumen, Gravel, Metal sheets, Shadows, Selfblocking Bricks, Meadows, and Bare soil. For these data, a set of 42776 labeled samples is available. In order to perform classification, we have randomly selected some samples for the pool of available labeled pixels



(a) (b)

Fig. 1: Pavia University image. (a) True color composition; (b) Available set of labeled samples with nine thematic classes (for color codes see [5]).

(a) (b) (c) (d)

Fig. 2: Example of features in the EMAP. (a) First principal component; Results of the attribute filtering obtained with the following attributes: (b) area; (c) moment of inertia; (d) standard deviation.

(with different rates) for making the dictionary and the training of the classifiers. The rest of the samples have been used for validation. A color composite and the map of the labeled samples is reported in Figure 1.

Two sets of features have been considered in the experiments:

- `Spec`: the spectral bands,
- `EMAP`: spatial features (144) obtained by an Extended Multi-Attribute Profile (EMAP) [10].

`Spec` is composed of the spectra as acquired by the sensor (i.e., the original image) whereas `EMAP` is a multilevel (e.g., multiscale) decomposition of the image obtained by morphological attribute filters. In greater details, a sequence of attribute filters (i.e., connected filters defined in the mathematical morphology framework) have been applied to the first four components of the Principal Component Analysis (PCA) considering several values of the filters (leading to progressively coarser images). Attribute filters process an image according to a given measure that we call attribute. Roughly speaking, the operator filters out regions that are not meeting a condition imposed by the filter (as that the area should be greater than a threshold) [10]. Four attribute have been considered in order to build the EMAP: area, length of the diagonal a region bounding box, moment of inertia (first moment invariant of Hu) and standard deviation of the values of the pixels in a region. An example of features in the EMAP is given in Figure 2.

`Spec` and `EMAP` features being different in nature show different characteristics. For example, `EMAP` can be seen as a mostly piece-wise smooth function in which samples might alternate intervals of constant or smooth values and punctual large variations. `Spec` values are in general smoother. However, samples belonging to different classes are more separable in the `EMAP` feature space than in the `Spec` one.

Eight collaborative-based representation classifiers have been considered in the experiments:

- $\mathtt{SRC}^+$: Collaborative classifier with ℓ_1 regularization (SRC) and NN constraint
- `SRC`: Collaborative classifier with ℓ_1 regularization (SRC)
- $\mathtt{CRC}^+$: Collaborative classifier with ℓ_2 regularization (CRC) and NN constraint
- `CRC`: Collaborative classifier with ℓ_2 regularization (CRC)
- $\mathtt{NSC}_1^+$: Collaborative classifier with ℓ_1 regularization and NN constraint
- $\mathtt{NSC}_1$: Collaborative classifier with ℓ_1 regularization
- $\mathtt{NSC}_2^+$: Collaborative classifier with ℓ_2 regularization and NN constraint
- $\mathtt{NSC}_2$: Collaborative classifier with ℓ_2 regularization.

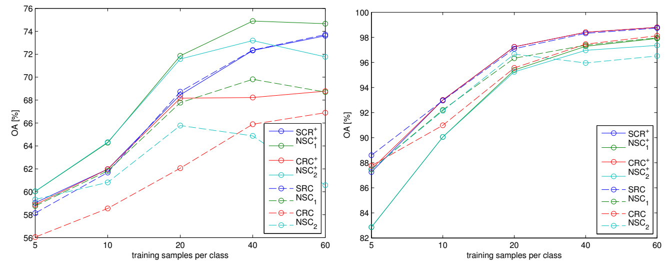
Four values of λ, the coefficient weighting the regularization term have been considered. Namely $\{0, 10^{-4}, 10^{-3}, 10^{-2}\}$. Five different sizes of the training set have been considered. Specifically the training was done with a fixed number of training samples per class in the interval $\{5, 10, 20, 40, 60\}$.

For every setting (i.e., classifier, λ and training set size) we have run the classification with 5 different random realizations of the training sets. All the results reported refers to the average of the values of Overall Accuracy (OA), i.e., percentage of pixels correctly classified, computed on each of the five realizations.

3.2. Experiments

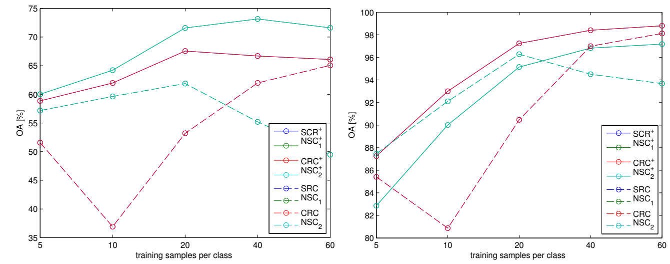
3.2.1. Comparison among classification strategies

We will at first compare the different classification strategies. In order to compare the results we have chosen to report the maximum of OA obtained over the experiments with the four values of λ. This is because, the different classification strategies react differently to the same regularization so selecting a certain value of λ might favor some strategies over the others. The results obtained are in Figure 3. When considering the



(a) (b)

Fig. 3: Maximum value of OA achieved over the different values of lambda $\lambda = \{0, 10^{-4}, 10^{-3}, 10^{-2}\}$. Overall accuracy reported for (a) `Spec` and (b) `EMAP` features for different training samples rates.



(a) (b)

Fig. 4: Influence of the NN constraint. Overall accuracy reported for (a) `Spec` and (b) `EMAP` features for $\lambda = 0$ for different number of training samples.

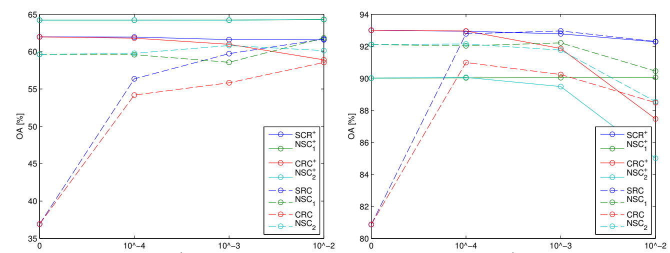
`Spec` features, the strategy of `NSC` with NN constraint have globally reached the greatest results in terms of OA. $\texttt{SRC}^+$, `SRC`, $\texttt{NSC}_1$ and $\texttt{CRC}^+$ perform closely even if for larger sets of training samples the SRC strategy results more effective. Approaches without NN constraint and with ℓ_2 regularization achieved the lowest performances showing the need of (stronger) regularizations for such type of features. Different conclusions can be drawn looking at the results from `EMAP`. All techniques achieved in general results within a range of about 3% showing a more stable behavior also for varying sizes of the training. This might be due to the easier discriminability of the samples in this feature space w.r.t. in the `Spec` space. $\texttt{SRC}^+$, `SRC` and $\texttt{CRC}^+$ achieved very similar results outperforming the other strategies. $\texttt{NSC}_1$ and $\texttt{NSC}_2$ outperformed the counterparts with NN constraint.

3.2.2. Influence of the NN constraint

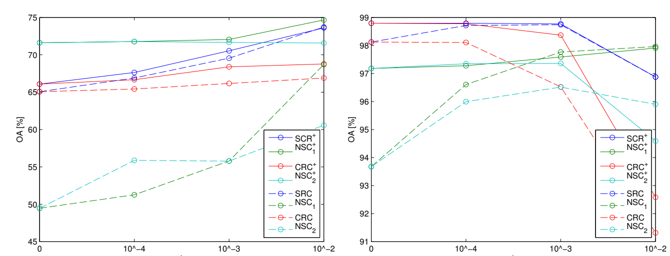
For this analysis we consider the results obtained by the different classification strategies without any other constraint on the reconstruction code (i.e., $\lambda = 0$). We recall that in this setting, SRC equals CRC and NSC_1 equals NSC_2 since the regularization term is inactive. The classification accuracies are reported in Figure 4. From the obtained results it is possible to state that in most of the cases the NN constraint leads to results that are superior with respect to those obtained by the same classifier without constraint. Considering the scenario with no regularization, the NN constraint helps in restraining the space of solutions limiting the possibility that samples of a class could be well represented by other classes. However, as shown in Figure 4(b) when the dictionary is small (i.e., few training samples available) the constraint can be too restrictive even affecting the performances.

3.2.3. Influence of the regularization

The results for a number of 10 and 60 training samples per class are reported in Figure 5. Globally, as also seen in the previous analysis, strategies with NN constraint attain results that are less affected by the weight of the regularization. Among all, $\texttt{NSC}_1^+$ has shown constant behavior for both feature types and training sample size. For strategies without the NN constraint, in general the regularization leads to an improvement of the results. However, too high values of λ can degrade the performances as seen for the `EMAP` features.



(a) `Spec` (b) `EMAP`



(c) `Spec` (d) `EMAP`

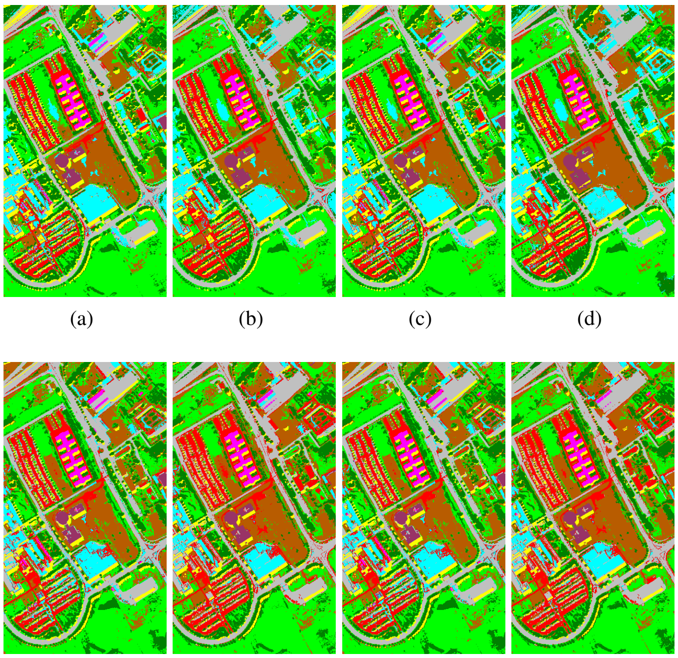
Fig. 5: Influence of the regularization. Overall accuracy reported for `Spec` and `EMAP` features with (a)-(b) 10 and (c)-(d) 60 training samples per class considering $\lambda = \{0, 10^{-4}, 10^{-3}, 10^{-2}\}$.

The timings are reported in Table 1 showing the significantly reduced complexity of schemes with ℓ_2 regularization and without NN constraint.

Examples of classification maps are shown in Figure 6.

λ	$\texttt{SRC}^+$	$\texttt{SRC}$	$\texttt{CRC}^+$	$\texttt{CRC}$	$\texttt{NSC}_1^+$	$\texttt{NSC}_1$	$\texttt{NSC}_2^+$	$\texttt{NSC}_2$
0	1093.1	8.9	941.9	8.20	562.6	8.2	536.0	8.3
10^{-4}	1152.9	452.6	951.9	9.1	691.5	472.8	542.7	8.9
10^{-3}	1139.6	829.9	942.5	9.5	673.6	513.2	509.3	8.1
10^{-2}	1106.6	948.1	569.7	9.1	693.8	527.9	300.6	8.7

Table 1: Timings in $[s]$, for $\texttt{EMAP}$ features with 60 samples per class and different values of λ.



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 6: Classification maps obtained for $\texttt{EMAP}$ features, $\lambda = 10^{-4}$ and 10 training samples per class, for the following classifiers (OA [%] values in brackets): (a) $\texttt{SRC}^+$ (92.07); (b) $\texttt{SRC}$ (91.76); (c) $\texttt{CRC}^+$ (91.75); (d) $\texttt{CRC}$ (91.01); (e) $\texttt{NSC}_1^+$ (87.20); (f) $\texttt{NSC}_1$ (90.10); (g) $\texttt{NSC}_2^+$ (87.21); (h) $\texttt{NSC}_2$ (89.62).

4. CONCLUSION

In this paper we have conducted a comparative study among several collaborative-based representation classification strategies. In addition to the classical SRC and CRC methods we have investigated the NSC with regularization, being a much less explored approach. The analysis aimed at showing the influence of the different parameters such as regularization terms and NN constraint in the classification results. From the results of the experiments we have reached some conclusions: i) SRC and CRC perform overall similarly, though the sparse regularization leads in general to more stable and slightly higher results; ii) the NN constraints is relevant in terms of classification, in general leading an increase in the performances expecially if other regularizations on the solutions are not present; iii) NSC with regularization have proven to be a competitive approach achieving results similar and in some cases superior to those of SRC and CRC; iv) techniques based on optimization endowed by a closed form solution (i.e., $\texttt{CRC}$ and $\texttt{NSC}_2$) achieve in general similar results, though with a greater variability w.r.t. the others. As future directions of this work, we plan to confirm experimentally the conclusions found here on other datasets and to better investigate the relationship between reconstruction error in the representation and classification accuracy.

REFERENCES

[1] Michael Zibulevsky and Michael Elad, “L1-l2 optimization in signal and image processing,” *Signal Processing Magazine, IEEE*, vol. 27, no. 3, pp. 76–88, 2010.

[2] John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry, and Yi Ma, “Robust face recognition via sparse representation,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, 2009.

[3] M-D Iordache, José M Bioucas-Dias, and Antonio Plaza, “Sparse unmixing of hyperspectral data,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 6, pp. 2014–2039, 2011.

[4] Yi Chen, Nasser M Nasrabadi, and Trac D Tran, “Hyperspectral image classification using dictionary-based sparse representation,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 10, pp. 3973–3985, 2011.

[5] Benqin Song, Jun Li, M. Dalla Mura, Peijun Li, A. Plaza, J.M. Bioucas-Dias, J. Atli Benediktsson, and J. Chanussot, “Remotely sensed image classification using sparse representations of morphological attribute profiles,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 5122–5136, Aug 2014.

[6] Q Sami ul Haq, Linmi Tao, Fuchun Sun, and Shiqiang Yang, “A fast and robust sparse approach for hyperspectral data classification using a few labeled samples,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 6, pp. 2287–2302, 2012.

[7] D Zhang, Meng Yang, and Xiangchu Feng, “Sparse representation or collaborative representation: Which helps face recognition?,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 471–478.

[8] J. Li, H. Zhang, L. Zhang, X. Huang, and L. Zhang, “Joint collaborative representation with multitask learning for hyperspectral image classification,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. PP, no. 99, pp. 1–14, 2014.

[9] Sen Jia, Linlin Shen, and Qingquan Li, “Gabor feature-based collaborative representation for hyperspectral imagery classification,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 53, no. 2, pp. 1118–1129, Feb 2015.

[10] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, “Extended profiles with morphological attribute filters for the analysis of hyperspectral data,” *International Journal of Remote Sensing*, vol. 31, no. 22, pp. 5975–5991, 2010.