Impact of Feature Extraction and Feature Selection Techniques on Extended Attribute Profile-based Hyperspectral Image Classification

Rania Zaatour, Sonia Bouzidi and Ezzeddine Zagrouba
Research team SIIVA, LIMTIC laboratory, Higher Institute of Computer Science (ISI), University of Tunis El Manar, 2 Street Abou Rayhane Bayrouni, 2080 Ariana, Tunisia
rania.zaatour@fst.utm.tn, sonia.bouzidi@insat.rnu.tn, ezzeddine.zagrouba@fsm.rnu.tn

Keywords: Dimensionality Reduction, Principal Component Analysis (PCA), Local Fisher Discriminant Analysis (LFDA), Independent Component Analysis-based Band Selection, Extended MultiAttribute Profile (EMAP), Hyperspectral Image Classification.

Abstract: Extended multiattribute profiles (EMAPs) were introduced as morphological profiles built on the features of a hyperspectral image extracted using Principal Component Analysis (PCA). In this paper, we propose to replace PCA with other dimensionality reduction techniques. First, we replace it with Local Fisher Discriminant Analysis (LFDA), a supervised locality preserving DR method. Second, we replace it with two band selection techniques: *ICAbs*, an Independent Component Analysis (ICA) based band selection, and its modified version that we propose in this article and which we are calling *mICAbs*. In the experimental part of this paper, we compare the accuracies of classifying the sparse representations of the EMAPs built on features obtained using each of the aforementioned dimensionality reduction techniques. Our experiments reveal that LFDA gives, amongst all, the best classification accuracies. Besides, our proposed modification gives comparable to higher accuracies.

1 INTRODUCTION

Hyperspectral remote sensing images (HSI) are data cubes formed out of hundreds of tight correlated spectral bands. This type of data offers rich spectral and spatial information permitting a better distinction of the objects contained in the acquired scene, which leads to higher classification accuracies.

Most of the traditional HSI classification techniques exploit only the spectral information provided by the image. But, given the richness of the information offered by HSIs, it is obvious that combining both spectral and spatial information delivers more accurate classification results.

A few years ago, Extended MultiAttribute Profiles (EMAP) (Dalla Mura et al., 2010) have been introduced as morphological profiles allowing the extraction of both spatial and spectral information from HSIs.

Since first introduced, EMAPs were tested in several HSI classification tasks through which it has been proven that they enable the detailed modeling of the structural information of an image's objects while preserving the geometrical information as well as the spectral information of the data (Dalla Mura et al., 2010; Dalla Mura et al., 2011; Song et al., 2014; Li et al., 2014). This modeling depends on the type of attributes used to compute the EMAPs and therefore no prior knowledge of the processed image is required.

This morphological profile is defined as the concatenation of a set of images generated after filtering the features of a reduced HSI. These features are obtained after applying a dimensionality reduction method, as a preliminary step, to the given HSI.

Dimensionality reduction techniques are destined to lower the processing time and complexity, by minimizing the redundancy and reducing the hundreds of bands, while assuring to keep most of the information required to guarantee the effectiveness of the task to perform. To do so, dimensionality reduction methods act in one of the following ways: (a) extracting features, or (b) selecting features.

Feature extraction methods decorrelate the HSI's information and eliminate its redundancy by projecting the original data on a new lower-dimensional feature space, and then selecting the first few relevant features that contain most of the information in the data. On the other hand, feature selection techniques select a subset from the original data cube. These methods reduce the highly-dimensional data without

Zaatour R., Bouzidi S. and Zagrouba E.
Impact of Feature Extraction and Feature Selection Techniques on Extended Attribute Profile-based Hyperspectral Image Classification.
DOI: 10.5220/0006171305790586
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 579-586
ISBN: 978-989-758-225-7

changing its features. Interestingly enough, some feature selection techniques relate to feature extraction methods to select the most relevant bands.

When it comes to reducing the HSI's dimensionality before building its EMAP, only feature extraction techniques were used in the literature. When first introduced, EMAP was defined as a morphological profile built upon the first few principle components (PC) generated by PC Analysis (PCA) (Dalla Mura et al., 2010). This feature extraction method is the most commonly used technique to reduce HSI before building their EMAP.

Other than PCA, in the literature, EMAPs were built on the features obtained using other feature extraction techniques such as Independent Components Analysis (ICA) (Dalla Mura et al., 2011), Discriminant Analysis Feature Extraction (DAFE), Decision Boundary Feature Extraction (DBFE), and Nonparametric Weighted Feature Extraction (NWFE) (Ghamisi et al., 2014).

The only way where feature selection techniques were used in the context of classifying HSIs using their EMAP, was to reduce the profile built using all the original bands of the given image. Hence, in this case, we are generating a huge profile and then we are selecting the most representative features that will enhance the classification process.

Some of the feature selection techniques used for this purpose, mostly based on genetic algorithms, can be found in (Pedergnana et al., 2013; Ghamisi and Benediktsson, 2015; Tuia et al., 2014).

In this paper, we are interested in the HSI dimensionality reduction prior to the generation of EMAPs. For this purpose, we are proposing the use of three dimensionality reduction techniques that, to the authors' knowledge, were never used in such context.

First, we propose to use Local Fisher Discriminant Analysis (LFDA) instead of the most common PCA. Since EMAPs are computed based on the connected components of an original or transformed HSI, and since LFDA accounts for locality, we believe that the use of this feature extraction method will help generating more representative EMAPs. Moreover, LFDA overcomes the limitations of PCA as (a) it is supervised and (b) it discards the assumption of a Gaussian distribution of the data. Consequently, we strongly believe that this dimensionality reduction technique would give better results than the traditional PCA.

Second, we investigate the efficiency of feature selection techniques to reduce the HSI dimensionality prior to the generation of EMAPs. We are particularly interested in a feature selection method that exploits the transformation matrix of a feature extraction technique, namely ICA. This ICA-based band selection technique, hereinafter referred to as *ICAbs*, selects the bands that contribute the most to the ICA transformation. To do so, ICAbs evaluates ICA's unmixing matrix and selects the bands having the highest absolute average coefficient.

Moreover, in this article, we propose a new feature selection technique consisting in a slightly modified version of ICAbs. This new method that we are calling *mICAbs*, like *modified* ICAbs, is the same as the original band selection technique, except the fact that each one of them uses its own criterion to decide of the bands to be selected.

All the aforementioned dimensionality reduction techniques will be compared in the same HSI classification context using a sparse representation classification framework. These methods will be evaluated according to the classification accuracies we get when using them.

The rest of the paper is structured as follows: first, in section 2, we describe how to build an EMAP and how to use it in a HSI classification task based on sparse representations. Section 3 recalls the bases of PCA and LFDA and theoretically compares them. Section 4 details the algorithm of the original ICA-based band selection and introduces our proposed modification. Section 5 describes the used HSI and performs a comparison of the studied dimensionality reduction techniques. Finally, section 6 concludes the article.

2 HSI CLASSIFICATION USING EMAP SPARSE REPRESENTATION

In this section, we represent the way an EMAP is generated, and we introduce the sparse classification framework we will be using in the experimental part.

2.1 Extended Multi-Attribute Profile (EMAP)

We can think of an EMAP as a cube of grayscale images resulting from the application of attribute filters to the connected components of every feature selected/extracted from a HSI.

Attribute filters proceed depending on whether an attribute A (e.g., area, standard deviation, moment of inertia) computed on a connected component C_i of the concerned feature meets a predefined condition on a threshold value λ (e.g., $A(C_i) > \lambda$). If $A(C_i)$ meets the condition, C_i is kept unaltered. Otherwise, it is merged to the nearest connected component having a lower (respectively greater) gray level, and this merging is called thinning γ (respectively thickening ϕ).

If we apply several thickenings and thinnings to the same feature f using a set of ordered thresholds $\{\lambda_1,...,\lambda_n\}$, we obtain an Attribute Profile (AP):

$$AP(f) = \{\phi_n(f),...,\phi_1(f), f, \gamma_1(f),...,\gamma_n(f)\} \quad (1)$$

Since a reduced HSI rh is a stack of r features, concatenating the APs generated from every feature f_i leads us to the definition of Extended Attribute Profile (EAP):

$$EAP(rh) = \{AP(f_1), AP(f_2),...,AP(f_r)\} \quad (2)$$

If we decide to use multiple attributes and generate an EAP based on each one of them, the concatenation of the generated profiles is the so called EMAP.

Fig. 1 summarizes the above mentioned steps of building an EMAP.

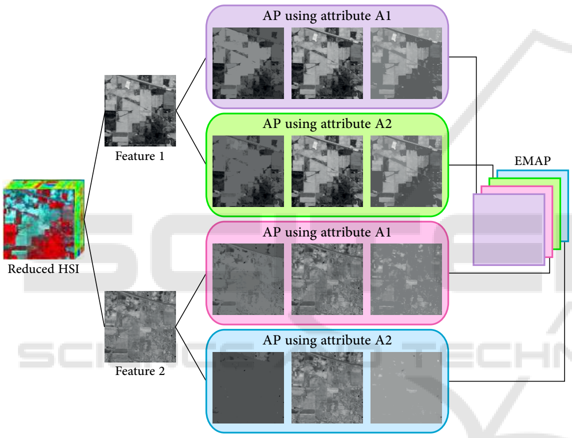

Figure 1: Simplified process of building an EMAP using a 2-feature reduced HSI and only 2 attributes.

2.2 Sparse Classification Framework

We chose to compare the studied DR techniques through a classification task using a sparse representation classification framework to better represent the highly dimensioned EMAPs.

Let $D \in \mathbb{R}^{a\times l}$ be a training dictionary made of a l-dimensional atoms representing c classes such that $D = \{d_1,...,d_a\} = \{D_1,...,D_c\}$, where D_k contains the a_k samples of class k, and $\sum_{k=1}^{c} a_k = a$. Considering that a signal y can be represented by a linear combination of D's atoms, we have :

$$\begin{aligned} y &\approx d_1x_1 + ... + d_ax_a \\ &= [d_1,...,d_a][x_1,...,x_a]^\top \\ &= Dx + \varepsilon \end{aligned} \quad (3)$$

where ε is the representation error and $x = [x_1,...,X_a]^\top = [X_1^\top,...,X_c^\top]^\top$ is a sparse a-dimensional vector whose every value X_i is the regression coefficients vector of the class i. If y belongs to class i, we assume that it is well approximated by D_iX_i (Song et al., 2014; Chen et al., 2011). Then, x is a sparse vector where $X_j = 0 \quad \forall j \neq i$.

To classify a signal y, we need to find an approximation of the sparse vector x subject to $y = Dx + \varepsilon$ by solving the following constrained optimization problem:

$$\min_x \frac{1}{2}\|y - Dx\|_2^2 + \tau\|x\|_1, \quad x \geq 0 \quad (4)$$

where τ is a Lagrange multiplier, which tends to 0 when ε tends to 0.

In our context, the dictionary's atoms are EMAPs of randomly selected samples from each class and y is the EMAP of the current pixel to be classified. The whole classification process consists in attributing to the current pixel, whose EMAP is y, the class label of the dictionary's atom that has the highest non-null coefficient in the approximated x.

3 PRINCIPAL COMPONENT ANALYSIS VS. LOCAL FISHER DISCRIMINANT ANALYSIS

Generally, the conventional feature extraction technique used to reduce HSIs is Principal Component Analysis (PCA) (Hotelling, 1933). This unsupervised second-order statistics-based method transforms the original bands, using a linear combination, into decorrelated and variance maximizing Principle Components (PC), then, the first few r PCs are selected. These extracted features, and since they represent most of the information that the original data offer, were exploited to build EMAPs, when this profile was introduced and in most cases where it was used afterward.

Nevertheless, in certain cases, when used prior to classification, PCA may fail to reduce the dimensions while keeping all the useful and representative information. This is due to the facts that (a) PCA assumes that class distribution is Gaussian although real-life data is more likely to have a complex multimodal distribution (Martínez and Kak, 2001), (b) it tends to omit some information that might be useful to the classification process (Prasad and Bruce, 2008) and (c) due to its unsupervised nature, PCA does not count for class labels and thus, might lead to a more complex class separability (Varghese et al., 2012).

A feature extraction technique that claims overcoming these limitations is the Local Fisher Discriminant Analysis (LFDA). It was introduced in (Sugiyama, 2007) as the combination of two well known dimensionality reduction techniques: Fisher Discriminant Analysis (FDA) (Fisher, 1936) and Locality Preserving Projection (LPP) (Niyogi, 2004).

LFDA can be thought of as a localized version of FDA where multimodal input data are handled effectively and the local structure of nearby samples in the original space is preserved (Li et al., 2012).

Let $X = \{x_1, x_2, ..., x_p\}, x_i \in \mathbb{R}^n$ be a set of p samples and $y_i \in \{1, 2, ..., c\}$ be the class labels, where c represents the total number of classes and p_l represents the number of samples of class l ($\sum_{l=1}^{c} p_l = p$). LFDA computes a *local* between-class scatter matrix S^{lb} and a *local* within-class scatter matrix S^{lw} defined respectively in (5) and (6).

$$S^{lb} = \frac{1}{2} \sum_{i,j=1}^{p} W_{i,j}^{lb} (x_i - x_j)(x_i - x_j)^\top \tag{5}$$

$$S^{lw} = \frac{1}{2} \sum_{i,j=1}^{p} W_{i,j}^{lw} (x_i - x_j)(x_i - x_j)^\top \tag{6}$$

where W^{lb} and W^{lw} are $p \times p$ matrices, respectively defined as in (7) and (8), using $A_{i,j}$ which is the $p \times p$ LPP affinity matrix that characterizes the distance between data samples.

$$W_{i,j}^{lb} = \begin{cases} A_{i,j}(\frac{1}{p} - \frac{1}{p_l}) & if \quad y_i = y_j = l, \\ \frac{1}{p} & if \quad y_i \neq y_j, \end{cases} \tag{7}$$

$$W_{i,j}^{lw} = \begin{cases} \frac{A_{i,j}}{p_l} & if \quad y_i = y_j = l, \\ 0 & if \quad y_i \neq y_j. \end{cases} \tag{8}$$

Using both the local between-class and the local within-class scatter matrices, defining LFDA's transformation matrix, T, is reduced to the maximization of an objective function, as shown in (9).

$$T = \arg\max[tr((T^\top S^{lw} T)^{-1} \;\; T^\top S^{lb} T)] \tag{9}$$

where $T \in \mathbb{R}^{n \times r}$ and r is the new reduced dimension.

4 ICA-BASED BAND SELECTION TECHNIQUES

In order to generate the final independent components (IC), IC Analysis (ICA) defines a transformation matrix that unmixes the original signal sources based on their statistical independency measured by mutual information (Wang and Chang, 2006).

Let S be the original set of mixed source signals. ICA aims at separating these signals in order to provide a new set of statistically independent sources X. To do so, ICA searches for an unmixing matrix U, such that:

$$X = US \tag{10}$$

When ICA is used as a HSI DR technique, S is an $n \times p$ matrix referring to the concerned HSI containing n bands and p pixels. On the other hand, X refers to the reduced $r \times p$ set of the resulting ICs. Therefore, U is an $r \times n$ matrix, where:

$$x_{ij} = \sum_{k=1}^{n} u_{ik} s_{kj}, \quad i = 1, ..., r, \quad j = 1, ..., p \tag{11}$$

If we consider that u_{ik} is the weight of information the k^{th} band is containing regarding to the i^{th} IC, $u_k = [u_{1k}, ..., u_{rk}]$ is then the vector of contributions of the k^{th} band to the ICA transformation. From this assumption, the ICA-based band selection technique we mentioned earlier in the introduction (Du et al., 2003), hereinafter referred to as *ICAbs*, looks for the bands having the higher average absolute weight $\bar{u_k}$ given in (12), and selects them for having the highest contribution to the ICA transformation.

$$\bar{u_k} = \frac{1}{r} \sum_{i=1}^{r} |u_{ik}| \tag{12}$$

In this paper, we are proposing a modified version of ICAbs. This technique that we will call *mICAbs*, like *modified* ICAbs, will follow the same steps as the original method. What we are changing is the criterion according to which we will be considering that a band is having a higher contribution than an other one. Instead of looking for the bands having the highest average absolute weight, we will look for those having the highest entropy.

This information theory concept is a statistical measure of randomness that is related to the amount of information a random variable is containing: the higher the entropy is, the larger the amount of information in the data is (Bajcsy and Groves, 2004). Our choice of the entropy is encouraged by the fact that ICA already uses it to define its ICs. Then, as long as we are concerned, we believe that the use of entropy will ameliorate the band selection process.

So, our proposed approach is to replace (12) by (13), where we calculate the entropy of the contribution vector u_k instead of its average absolute weight.

$$Eu_k = Entropy(u_k) = -\sum_{i=1}^{r} p(u_{ik}) \; \log \; p(u_{ik}) \tag{13}$$

where $p(u_{ik})$ represents the mass probability of an event u_{ik} from a finite set of possible values (Martínez-Usó et al., 2007) .

Once the entropy of every contribution vector u_k is computed, we obtain a sequence of band contribution's entropies defined in (14).

$$[Eu_1, ..., Eu_k, ..., Eu_n] \tag{14}$$

Each Eu_k of this sequence tells how informative the contribution of the k^{th} band to ICA's transformation matrix is. Hence, selecting the bands having the same indexes as those of the highest elements in this sequence is equivalent to selecting the bands that contribute the most to the ICA's transformation.

5 EXPERIMENTAL RESULTS

This experimental section is organized as follows. First, we compare the accuracies of classifying EMAPs built upon features extracted by LFDA to the ones obtained using features extracted by PCA. Then, we compare the accuracies of classifying EMAPs built upon bands selected by ICAbs to the ones obtained using bands selected by mICAbs.

In our context, EMAPs are built using the reduced data which is rescaled to the range $[0,255]$ and converted to integers before applying the AFs. The employed attributes are (a) area, which measure the size of regions, using threshold values ranging from 50 to 500 with a stepwise increment of 50, and (b) standard deviation, which measures the homogeneity of the pixels in every region, using threshold values ranging from 2.5% to 20% with a stepwise increment of 2.5%.

The dictionary is made of the EMAPs of 10% of every class' total samples. In order to guarantee a maximum of equal chances to all compared methods, all the dictionaries used in this experimental part are made of the EMAPs of the same randomly selected pixels.

In the literature, it has been proven that Sparse UNmixing by variable Splitting and Augmented Lagrangian (SUnSAL) is more efficient at solving equation (4) than other classifiers (Song et al., 2014). For this reason, and since our aim is to compare the DR methods and not the classification process, we will be using only SUnSAL for our experiments.

The HSI we used to test our approach is the well known hyperspectral image captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over the Indian Pines region situated in the Northwest of Indiana, USA, in june 1992. This image, known as AVIRIS Indian Pines[1], is composed of 145×145 pixels and 220 spectral bands which have been reduced to 200 bands in order to avoid both noise and water absorption phenomena. Fig. 2 shows the AVIRIS Indian Pines' ground truth with it's corresponding color legend.

[1]Available online at https://purr.purdue.edu/publications/1947/1

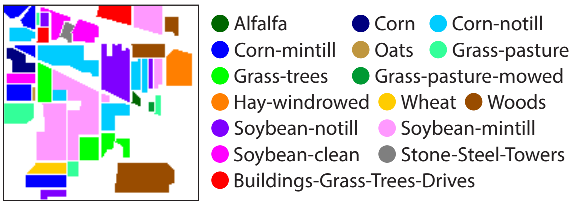

Figure 2: Ground truth of the AVIRIS Indian Pines.

In order to evaluate the effectiveness of the classification process we used the following metrics: the average accuracy (AA), the overall accuracy (OA) and the average reliability (AR).

5.1 Comparison of Classification Results using LFDA and PCA

Fig. 3 illustrates the AAs we got using PCA and LFDA, according to the variation of the number of extracted features.

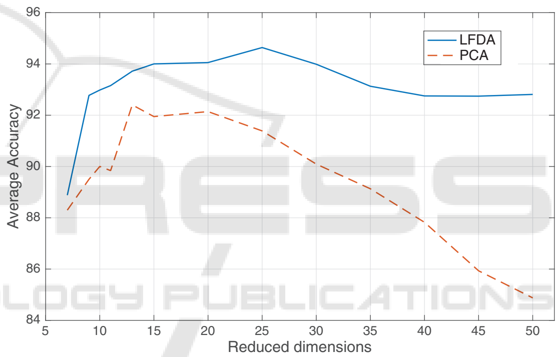

Figure 3: Average accuracy (AA) using LFDA and PCA according to DR variation.

From Fig. 3, we can deduce that LFDA outperforms PCA in all the tested cases. The gap between the AAs of both techniques rises as the number of kept features rises, and reaches up to 7.9% when EMAPs are built upon 50 features.

Furthermore, we can remark that LFDA is less sensitive to the curse of dimensionality. This curse, also known as Hughes Phenomenon (Hughes, 1968), consists in the fact that the higher the number of features is, the lower the classification accuracy is.

In Fig. 4, we can deduce that both feature extraction techniques give very close OAs when using a small number of features. Besides, PCA outperforms LFDA in some cases. But, starting from 20 features, LFDA mostly keeps its high accuracies while the OAs obtained using PCA decline.

When using a small number of features, we can remark that the AAs and OAs give opposite behaviors. In fact, when LFDA gives higher AA, PCA gives

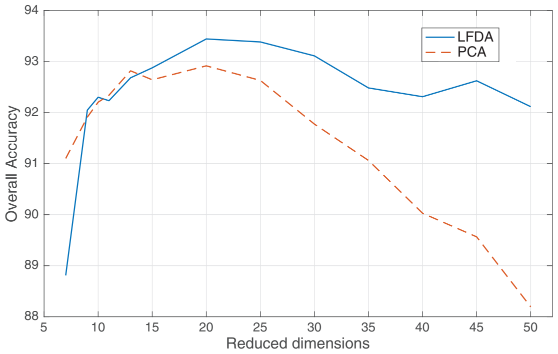

Figure 4: Overall accuracy (OA) using LFDA and PCA according to DR variation.

higher OA. This can be explained by the fact that PCA gives better representations of larger classes, and so gives a better OA, while LFDA gives a good representation of both small and large areas, and that is why it has a better AA.

According to Fig. 5, LFDA is more reliable than PCA in all the tested cases.

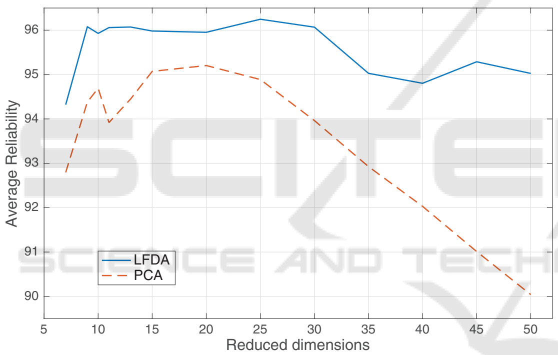

Figure 5: Average reliability (AR) using LFDA and PCA according to DR variation.

To conclude this part of the experimentation, we can assert that LFDA enhances the accuracy of EMAP-based HSI classification, when used, instead of the most commonly used PCA, to reduce HSI's dimensionality prior to generating its EMAP.

5.2 Comparison of Classification Results using ICA-based Band Selection Techniques

In this part of the experimentation, we evaluate and compare ICAbs and our proposed mICAbs as feature selection techniques used to reduce HSI's dimensionality prior to the generation of EMAPs.

We can already assume that the results we will be getting using feature selection techniques would not be as accurate as the ones obtained with the feature extraction ones. This is due to the fact that feature extraction transformations exploit the whole original data in order to extract the most pertinent features. However, feature selection techniques select the bands as they are without altering them or consolidating them.

Fig. 6 illustrates the AAs obtained using both ICA-based band selection techniques. We can deduce that both methods give comparable accuracies, with mICAbs giving better results in most of the tested cases. This improvement is measured by a difference reaching up to 4.4%.

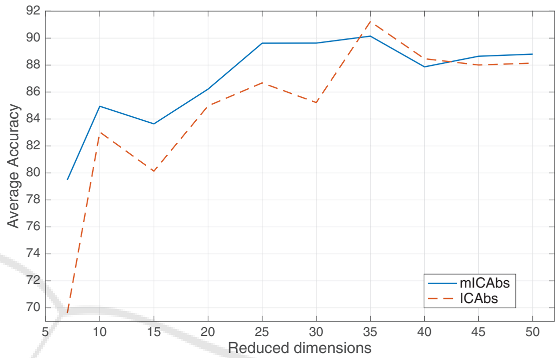

Figure 6: Average accuracy (AA) using ICAbs and mICAbs according to DR variation.

The same conclusion can be made from comparing the OAs, shown in Fig. 7, where mICAbs enhances the accuracy by a difference reaching up to 6.8%,

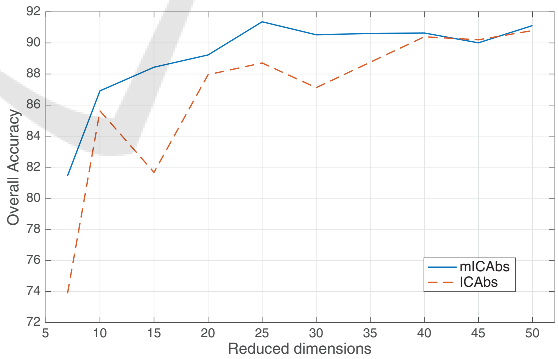

Figure 7: Average accuracy (OA) using ICAbs and mICAbs according to DR variation.

Moreover, Fig. 8 confirms one more time that both ICA-based band selection techniques give very close results with mICAbs outperforming ICAbs in most of the studied cases.

Besides, we can conclude that with both feature selection techniques, unlike with feature extraction ones, the classification is not affected by the curse of dimensionality. That is to say, we do not notice a high

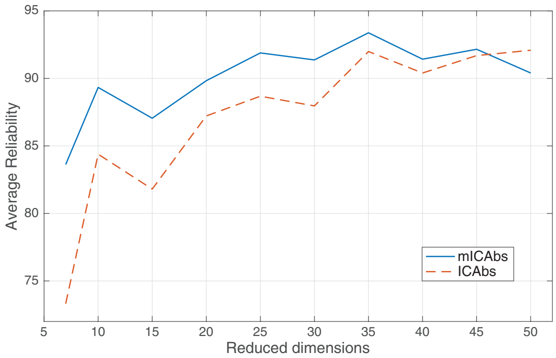

Figure 8: Average accuracy (AR) using ICAbs and mICAbs according to DR variation.

decreasing in accuracies when using a higher number of features (selected bands in this case).

This might be explained by the fact that, when using feature selection techniques, we are selecting the most representative bands from a highly correlated and redundant data. Thus, we are selecting bands that are not necessarily very similar, and consequently, we are minimizing the redundancy. Instead, in the feature extraction context, we are generating features that accumulate most of the information contained in the original data. Then, we will have a high redundancy since we are trying to put the maximum of information in each feature.

It is worth noting that the obtained accuracies show a lot of tops and bottoms, unlike the ones we got in the feature extraction case, which tend to be monotonous. In fact, in every application, LFDA/PCA apply the same transformation to the original data and rank the features according to the information they hold. Hence, using a bigger number of reduced dimensions is equivalent to keeping the old features and adding new ones so that the new set holds even more information.

In contrast, the feature selection techniques we used here are based on ICA which, when defining its unmixing matrix, starts with random initial projection vectors. Due to this randomness, the ICA transformation is not the same at every run. Therefore, the mICAbs to ICAbs comparisons made above are not based on the *same* unmixing matrix.

In fact, at every execution, a new ICA transformation matrix is defined and then the selection process is made according to this matrix. Consequently, when selecting bands from the Indian Pines, whether using ICAbs or mICAbs, we will not be getting the same exact set of bands at every run neither, therefore, the same accuracy. This latter may rise or fall according to the used ICA transformation matrix, which can explain the changing accuracies we had in the last three figures.

6 CONCLUSION

In this paper, we proposed to replace PCA, the common feature extraction technique used to reduce a HSI's dimensionality before generating its EMAP, by other dimensionality reduction techniques.

On one hand, as a feature extraction technique, we proposed the use of LFDA which proved to be, according to the experimental results, better than PCA. This might be explained by the fact that LFDA takes into consideration both the locality and the class labels in order to maximize the between-class variance while minimizing the within-class variance. Thus, LFDA gives a reduced version of the image where classes are easily discriminated.

On the other hand, we explored the effectiveness of some feature selection techniques in the same context. In fact, we proposed to replace PCA with two ICA-based band selection methods: ICAbs, which is existent in the literature, and mICAbs, our proposed modified version of ICAbs.

Both feature selection techniques are based on the unmixing matrix of ICA. The only difference between them is the criterion according to which bands are selected: ICAbs is based on the average absolute contribution and mICAbs is based on the entropy of contributions.

Both used feature selection techniques proved to give close results that are not as good as those we get when using feature extraction techniques. Moreover, our proposed mICAbs proved to enhance the classification accuracy, compared to the original ICAbs.

ACKNOWLEDGEMENT

The authors would like to thank Mr. Mauro Dalla Mura for providing the AP implementation.

REFERENCES

Bajcsy, P. and Groves, P. (2004). Methodology for hyperspectral band selection. *Photogrammetric Engineering & Remote Sensing*, 70(7):793–802.

Chen, Y., Nasrabadi, N. M., and Tran, T. D. (2011). Hyperspectral image classification using dictionary-based sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10):3973–3985.

Dalla Mura, M., Benediktsson, J. A., Waske, B., and Bruzzone, L. (2010). Extended profiles with morphological attribute filters for the analysis of hyperspectral data. *International Journal of Remote Sensing*, 31(22):5975–5991.

Dalla Mura, M., Villa, A., Benediktsson, J. A., Chanussot, J., and Bruzzone, L. (2011). Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. *IEEE Geoscience and Remote Sensing Letters*, 8(3):542–546.

Du, H., Qi, H., Wang, X., Ramanath, R., and Snyder, W. E. (2003). Band selection using independent component analysis for hyperspectral image processing. In *Applied Imagery Pattern Recognition Workshop, 2003. Proceedings. 32nd*, pages 93–98. IEEE.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188.

Ghamisi, P. and Benediktsson, J. A. (2015). Feature selection based on hybridization of genetic algorithm and particle swarm optimization. *IEEE Geoscience and Remote Sensing Letters*, 12(2):309–313.

Ghamisi, P., Benediktsson, J. A., and Sveinsson, J. R. (2014). Automatic spectral–spatial classification framework based on attribute profiles and supervised feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5771–5782.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24.

Hughes, G. P. (1968). On the mean accuracy of statistical pattern recognizers. *Information Theory, IEEE Transactions on*, 14(1):55–63.

Li, J., Zhang, H., and Zhang, L. (2014). Supervised segmentation of very high resolution images by the use of extended morphological attribute profiles and a sparse transform. *IEEE Geoscience and Remote Sensing Letters*, 11(8):1409–1413.

Li, W., Prasad, S., Fowler, J. E., and Bruce, L. M. (2012). Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4):1185–1198.

Martínez, A. M. and Kak, A. C. (2001). Pca versus lda. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(2):228–233.

Martínez-Usó, A., Pla, F., Sotoca, J. M., and García-Sevilla, P. (2007). Clustering-based hyperspectral band selection using information measures. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):4158–4171.

Niyogi, X. (2004). Locality preserving projections. In *Neural information processing systems*, volume 16, page 153. MIT.

Pedergnana, M., Marpu, P. R., Dalla Mura, M., Benediktsson, J. A., and Bruzzone, L. (2013). A novel technique for optimal feature selection in attribute profiles based on genetic algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 51(6):3514–3528.

Prasad, S. and Bruce, L. M. (2008). Limitations of principal components analysis for hyperspectral target recognition. *IEEE Geoscience and Remote Sensing Letters*, 5(4):625–629.

Song, B., Li, J., Dalla Mura, M., Li, P., Plaza, A., Bioucas-Dias, J. M., Benediktsson, J. A., and Chanussot, J. (2014). Remotely sensed image classification using sparse representations of morphological attribute profiles. *IEEE transactions on geoscience and remote sensing*, 52(8):5122–5136.

Sugiyama, M. (2007). Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *The Journal of Machine Learning Research*, 8:1027–1061.

Tuia, D., Volpi, M., Dalla Mura, M., Rakotomamonjy, A., and Flamary, R. (2014). Automatic feature learning for spatio-spectral image classification with sparse svm. *IEEE Transactions on Geoscience and Remote Sensing*, 52(10):6062–6074.

Varghese, N., Verghese, V., Gayathri, P., and Jaisankar, N. (2012). A survey of dimensionality reduction and classification methods. *International Journal of Computer Science and Engineering Survey*, 3(3):45.

Wang, J. and Chang, C.-I. (2006). Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE transactions on geoscience and remote sensing*, 44(6):1586–1600.