A new feature extraction approach based on non linear source separation

Hela Elmannai[1], Mohamed Saber Naceur[2], Mohamed Anis Loghmari[3], Abeer AlGarni[4]
[1,4]Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Saudi Arabia
[2,3]Laboratoire de teledetection et remote sensing LTSIRS, ENIT, Tunisia

Article Info

ABSTRACT

A new feature extraction approach is proposed in this paper to improve the classification performance in remotely sensed data. The proposed method is based on a primary sources subset (PSS) obtained by nonlinear transform that provides lower space for land pattern recognition. First, the underlying sources are approximated using multilayer neural networks. Given that, Bayesian inferences update unknown sources' knowledge and model parameters with information's data. Then, a source dimension minimizing technique is adopted to provide more efficient land cover description. The support vector machine (SVM) scheme is developed by using feature extraction. The experimental results on real multispectral imagery demonstrates that the proposed approach ensures efficient feature extraction by using several descriptors for texture identification and multiscale analysis. In a pixel based approach, the reduced PSS space improved the overall classification accuracy by 13% and reaches 82%. Using texture and multi resolution descriptors, the overall accuracy is 75.87% for the original observations, while using the reduced source space the overall accuracy reaches 81.67% when using jointly wavelet and Gabor transform and 86.67% when using Gabor transform. Thus, the source space enhanced the feature extraction process and allow more land use discrimination than the multispectral observations.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Hela Elmannai
Department of Information Technology
College of Computer and Information Sciences
Princess Nourah bint Abdulrahman University
84428 Riyadh, Saudi Arabia
Email: hselmannai@pnu.edu.sa, hela.elmannai@gmail.com

1. INTRODUCTION

Providing suitable presentation is a challenging but important task for remote sensing and earth observation applications. Various data processing fields of remote sensing imagery have received an important attention due to the increasing development of multichannel sensors. Feature extraction approaches for multispectral and remotely data classsifcation is within the principal approaches. The Table 1 resume the feature extraction approaches for remote sensed image classification.

Existant works are based on spectral and spatial metrics [1]-[4], handcrafted feature extraction [1], [5] and semantic deep feature extraction [6]-[8]. Combining various feature extraction models has also been investigated to emprove the data representation efficieny [1], [4], [8]. In contrast to convencional methods that emply spectral or spatial data representation for land cover identification, a significant advantage of

source separartion is to allow a new representation for the orogonal data with less correlated component. The ability of source separation was not been investigated as feature extraction approach. In thid context, this work aims to study the source separation contribution in feature extraction for land cover classification.

Table 1. Features extraction approaches for land cover classification

Spatial metrics and Texture measures for land cover objects classification [1]
Parcels geometrical attributes including shape, height, proximity to major roads, similarity to neighbors [2]
Spectral indices (NDVI, MNDWI, NDBI) [3]
Object-based feature extraction based on spatial and spectral statistics [4]
High dimensional feature vector combining focal textures statistics (median, mean, and standard deviation) and Gray Level Co-occurrence Matrix derived in different kernel sizes [5]
Semantic features extraction using different deep convolutional neural networks models [6]
High-level semantic features extraction based on transfer learning and the Inception-ResNet-v2 model [7]
Deep semantic feature extraction using different models (VGG-S , VGG-M, VGG-F, VGG-VD16, VGG-VD16) [8]
Combining deep semantic features, spectral features and GLCM texture features [9]

Vision-based technology have been widely used for object of interest detection. For instance, in [10] the detecting task is based on histogram equalization and morphological processings. The method aimed to detect, classify and track road distresses. The proposed method reults include false alarms and require enhancent. To recognize hand motion, the method proposed in [11] is utilized to search region of interest, and applying the standard particle filter for motion recognition. In this research, we aim to use the source separation as a feature extraction method and to use the new retrieved sources for the classification task. Using vision-based technology can be based on the source space and can be compared to the results obtained when using the initial representation. Moreover, although machine learning and deep learning have simplified the classification task complexity, source separation may provide a new initial space for the learning phase and can be more accurate than initial observation.

Source separation aims to recover revealing hidden factors that generate measured signals. In this framework, the blind source separation process is of considerable importance for such dataset [12]. Considering linear mixture, many authors have proposed interesting works based on BSS with considering of different assumptions and simplifications. Linear mixture models have interested many researches. They assume the pixel as a linear mixture of the radiances reflected by the end-members. Source's separation criteria represent the key assumption in the leaded separation algorithm. Some method are are based based on misfit function called contrast other methods are based on high static range. Other types of methods are based on algebraic criteria for second order statistics or non stationary source criteria. Another methods category is based on source time-frequency diversity and sparcity. The principal component analysis (PCA) is based on creating uncorrelated variables that maximize the variance. Independent component analysis (ICA) assumes that the new variables are non-gaussian and are statistically independent [13], [14]. However, many reflections occur to the radiances emitted by the soil which represent a mixture of materials. The mixture is therefore nonlinear. The non-linear BSS method aims to determine the original sources [15]-[17]. The main contributions of this paper are:

− Using nonlinear source separation for feature extraction: At our knowledge there is not existing approaches that have used the source separation for feature extraction. Majority of works use spectral data or feature extraction based on the observation data
− A quality guided dimension reduction based on the classification efficiency. The primary source are the most representative for the landcover categories.

In the next part, we will detail the methodology for multispectral image feature extraction. The feature extraction method based on nonlinear source separation and a qualitative dimension reduction will be detailed. We will present also the feature vectors generation and the SVM classifier. Experimentations results are provided and discussed in section 3. The paper ends with conclusion and future work.

2. RESEARCH METHOD

In order to build a feature extraction method for land pattern for remote sensing images, we propose the following steps:
− Feature extraction based on nonlinear source separation
− Dimension reduction
− Pixelwise classification and patchwise classification
− Evaluation of o the source space efficiency in the classification process compared to the original observations

The remote sensing classification method is detailed in Figure 1. Remote sensed collected signals may cause many manifolds due to nonlinear mixture, so the nonlinear source separation is needed to provide uncorrelated sources. Then we will reduce the source number using a qualitative criterion. After that, we will study the primary source contribution in a pixelwise classification and a patchwise classification. We will generate the learning data set containing homogeneous region for selected land cover thematic classes. The patchwise classification is based on learning classification.
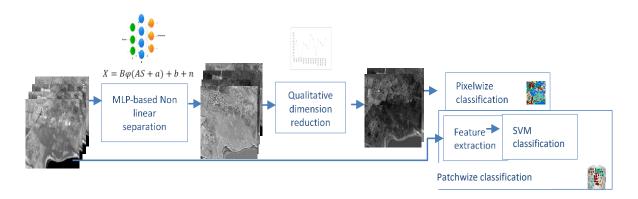

Figure 1. Methodology flowchart

2.1. Feature extraction based on nonlinear separation

The aim of this stage is to provide an efficient and compact presentation for pattern recognition. The process starts by a non linear separation that provides uncorrelated sources and approximated the nonlinear mixing phenomenon. The source set will be reduced to obtain a subset that presents efficiently the land cover. Then pattern transformations will generate feature vectors. The next parts will detail the feature extraction processes.

2.1.1. Non linear blind source separation based on bayesian inferenences

In our model, the separation is based on the nonlinear counterpart of PCA [16] to approximate the real mixture phenomenon. The mapping function contains a noise factor and the sources are gaussian mixture. The separation is performed by an MLP where the gradient descent tends to reduce the mutual function [17], [18]. Bayesian inferences estimate the mapping parameters from their priors. In (1) [19] represents the mapping function where $X(t)$ is a memoryless mixture of the sources $S(t)$ which are statistically independent. Estimated sources $Y(t)$ obtained by the separation model F as presented in (2).

$$X(t) = f(S(t)) + n(t) \text{ where } f : \Re^M \to \Re^N \tag{1}$$

$$Y(t) = F(X(t)) \tag{2}$$

The hidden layer activation function is the hyperbolic tangent [19], [20]. The unknown parameters are approximated iteratively using ensemble learning. Nonlinear mapping is presented in (3) where f is the nonlinear mixture function. The hidden layers contains H neurons. $A \in \Re^{NxH}$ and $B \in \Re^{NxH}$ are weight matrices. $a \in \Re^H$ and $b \in \Re^N$ are the bias vectors for repectively hidden and output layers. The unknown parameters, denoted Θ [21], have normal distribution $N(\bar{\theta}, \breve{\theta})$ where $\bar{\theta}$ is the mean and $\tilde{\theta}$ is the logarithm standard deviation. The posterior probability density (PDF) of unknwn paramteres are approximated [22] during the leaning by minimizing the Kullback-Leibler (KL) divergence between current PDFs $p(\theta|X)$ and their approximation $q(\theta|X)$. The factorial form presented in (4) is obtained thanks to the a-priori independency assumption.

$$f(S(t)) = B \tanh(A.S(t) + a) + b + n(t) \tag{3}$$

$$q(\theta, X) = \prod_i q(\theta_i, X) \tag{4}$$

The distributions misfit, measured with the KL divergence, is given by (5) [23]. We add the term $-\log(p(X)) = -\int p(X, \theta)d\theta$ to avoid the model evidence term calculation ($p(X)$). The divergence measure becomes as detailed in (6) [21]. The resulting mapping of the sources to the observations is presented in (7). As shown in (8) and (9) detail th unknown parameters distributions. Ones sources are approximated, the reduction algorithm performed to the source set is detailed in the next subsection.

$$KL\left(q(S,\theta)\|p(S,\theta|X)\right) = \int q(S,\theta)\log(\frac{q(S,\theta)}{p(S,\theta|X)})d\theta$$

(5)

$$D(\theta) = KL\left(q(\theta|X)|p(\theta|X)\right) - \log(p(X)) = D_q + D_p$$

(6)

$$X(t) \approx N(f(S(t)), e^{2\tilde{X}})$$

(7)

$$s(t) \approx N(0, e^{2\tilde{s}})$$

(8)

$$A \approx N(0,1) \quad B \approx N(0, e^{2\tilde{B}}) \quad a \approx N(\overline{a}, e^{2\tilde{a}}) \quad b \approx N(\overline{b}, e^{2\tilde{b}})$$

(9)

2.2. Source space dimension reduction and exemplars base generation

The dimension reduction process is an indispensable part for compact data presentation, data visualization or data classification preprocessing. The dimension reduction must preserve the required information existing in the initial signals [23]. Many methods have been developed [24]-[27]. They include data projections and transformations [28]-[30]. PCA method provides linear approximations for given data by encoding second order dependencies and finding the directions of maximal variance. Kernel PCA computes the principle components in the space produced by non linear mapping [31]. ICA method searches for independent components by rotations. Locally linear embedding [32] approach addresses the non linear dimensionality reduction. These mentioned methods lie to the feature extraction category that projects original data onto lower dimensional subspace. Other way for dimension reduction is feature selection and is based on finding a suitable subset form original data.

In this work, we present a dimension reduction method that refers to feature selection in the new presentation space of sources. The aim of the qualitative approach is to reduce the number of sources by distinguishing those describing better the land covers. The reason that leads the source space dimension reduction is the observations misclassifications due to band correlation, pixels scene heterogeneity and signal distortions by the atmosphere. Starting from uncorrelated sources obtained by the nonlinear separation, we aim to identify "primary" sources that ensure the best classification accuracy. An iterative process generates all source combination. Then for a well known land cover samples, the supervised classification will be performed for all source combination set. The best classification accuracy among all sources set determinates the PSS. This combination gives therefore the better description for the land covers. Sources that not appear in the SPP are called secondary sources subset (SSS). Those can be interpreted as hyper-information and can be used for further work. Compared to observation space, PSS space should give better classification accuracy and should describe more efficiently the land covers which will be proven in experimentation. Let consider the case of 4 observations. There are 4 obtained sources given by the Source Space=$\{S_1, S_2, S_3, S_4\}$. The possible source combinations are given by the set $Cs=\{\ \{S_1\},\ \{S_2\},\{S_3\},\{S_4\},\{S_1,\ S_2\},\{S_1,\ S_3\},\{S_1, S_4\},\{S_2,\ S_3\},\{S_2,\ S_4\},\{S_3,\ S_4\},\{S_1, S_2, S_3\},\{S_1, S_2, S_4\},\ \{S_1, S_3, S_4\},\{S_2, S_3, S_4\},\ \{S_1, S_2, S_3, S_4\}\ \}$. For these sets, the supervised classification is operated and accuracies are performed. The union of PSS and SSS constitutes the source space as presented in (10). PSS are determined by (11) where C denotes the source combination set.

$$S = PSS\bigcup SSS$$

(10)

$$PSS = \underset{C \in C_s}{Arg\ \max}(Classification\quad Accuracy)$$

(11)

The correspondent dimension reduction can be described as described in Algorihm 1. The notation $Comp(Y, X)$ where $Y \subset X$ denotes the complement of the subset Y in the set X. The reduction algorithm is performed by nested loops. $Accu$ denotes the classification accuracy for a given image set. $Accu(.)$ is a function that returns the accuracy of an image set throw a supervised classification. C_i^j is the combination set generated for i sources and providing j-combination sets. M is the source number. Acc_{curr} is the current

accuracy computed form the current source combination in the iterative process. Initiation is based on considering that all sources are primary $PSS=\{S_1, S_2, S_3, S_{4\}}$ and the maximum accuracy $MaxAcc$ is given by $Accu(C_M^M)$. The iterative process generates every primary source combination and computes the current accuracy $Accu_{curr}$ given by $Accu(C_i^j)$. If the $Accu_{curr}$ is greater than $MaxAcc$, the $MaxAcc$ will be updated by the value of $Accu_{curr}$, the current set C_i^j will be affected to the PSS, the SSS is the correspondent complementary in the source set.

Algorithm 1. Dimension reduction algorithm

```
Initation: MaxAcc = Accu(C_M^M)
S = {S1, S2, S3, S4}, PSS = C_M^M, SSS = φ
For i from 1 to M do
    For j from 1 to i do
        1- Generate the combination C_i^j ∈ C ; C = U_{i=1,M;j=1,M;j≤i} C_i^j
        2- Minimum Distance classification for C_i^j set |
        3- Current accuracy computation Accu_curr = Accu(C_i^j)
        4- If(MaxAcc < Accu_curr) Then
                        MaxAcc = Accu_curr
                        PSS = C_i^j
                        SSS = Comp(PSS,S)
            End if
    End For
End For
```

2.3. Classification method
2.3.1. Feature transformations

Image classification is a challenging problem in remote sensing applications namely agricultural researches, inundation and fire detection [33], [34]. Traditional classifiers are inefficient due to the correlation between observation channels. Moreover, observations are mixture of the original data. Thus, the non linear separation pre-processing is primordial for the classification task. Obtained approximated and decorrelated sources present better data representation compared to the original observations. Classificatin process may take advantages from the obtained sources. A deeper analysis of the obtained sources will allow a narrow relation between the source presentation and land pattern. PSS have lower dimension which reduce the classification complexity. In this work, we considerate macro patterns that present a set of land category such as urban areas, wetlands, parcels and lakes. The classification process is based on the new data presentations. Many descriptors were used for pattern recognition. The proposed fusion model concerns feature level which could improve the pattern recognition reliability [35]. In fact, feature fusion will improve the land categories classification process trough providing multi-feature representation. In our work, feature vector contains Gabor and Haar transformations [36]. Wavelets give a multi-description for regions in different scales and orientations which allows to represent the image singularities [37]. Low pass and high pass filters in each level of decomposition deconmposes the image into low and high frequencies sub-channels [35] which allows both an image approximation and high and lower frequencies identification. Wavelets basic functions are $\psi_{j,k=}\ 2^{j/2}\ \psi(2^j t\text{-}k)$ where ψ is the mother wavelet. For texture characterization, Gabor filters act as a multichannel filters offeroffers multichannels filters in different orientation u and scale v as defined in (12) [38]. The frequently used values for u and v are The frequekty used values are $v\in\{0,1,2,3,4\}$ and $u\in\{0,..,7\}$. For a given image I, the Gabor wavelet transform is given in (13) where * denotes the convolution operator [34].

$$\Psi_{u,v}(z) = \frac{\left\|k_{u,v}\right\|}{\sigma^2}\exp\left(-\frac{\left\|k_{u,v}\right\|\left\|z^2\right\|}{\sigma^2}\right)\left(\exp(ik_{u,v}z)-\exp(-\frac{\sigma^2}{2})\right) \tag{12}$$

$$J(z) = I(z) * \Psi_k(z) \tag{13}$$

2.3.2. The SVM classifier

Recent works were concentrated on developing machine learning methods for classification and recognition tasks. Support vector machines (SVM) method is used in this field and has proved his effectiveness by giving better classification accuracy than classic classifiers. SVM are widely used in land use and land cover classification and have the advantage of being particularly efficient for a huge feature space dimension. The classifier is based on genenerating a decision function defined by a set of learning instances [39]. The vector subset called support vectors (SV), determine a separating hyperplane separability.

For a binary classification, $V= \{v_1, \ldots, v_K\}$, $v_i \epsilon \mathcal{R}^d$ is the learning data and $c_i \in \{-1, +1\}$ is the label. The training set are the K training pairs $\{\{v_1, c_1\}, \ldots, \{v_K, c_K\}\}$. The hyperplane is defined by $w.v+b=0$ where the parameters w and b are constrained by $min_i|w.v_i + b| = 1$. The support vectors lie on two hyperplanes that are parallel to the optimal hyperplane defined by $w.v+b=0$. w and b are constrained by $min_i|w.v_i + b| = 1$. Maximazing the hyperplans margin leads to an optimization problem under constraints as presnted in (14) [34] with $\sum_{i=1}^{K} a_i c_i = 0$ and $0 \leq a_i \leq \zeta$, where ζ is a penalty factor. The optimal solution α*determinates the support vector set SV.

$$\alpha^* = \arg \min_\alpha \left(\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \alpha_i \alpha_j c_i c_j \langle v_i, v_j \rangle - \sum_{l=1}^{k} \alpha_l \right) \qquad (14)$$

The optimal hyperplane is given by $w^* = \sum_{i=1}^{K} a_i c_i v_i$ and $b^* = -\frac{1}{2} \langle w^*, v_p + v_q \rangle$ where α_i are the Lagrange multipliers and v_p and v_q are any support vector from each class. The classifier function is given by the funtion $X(v) = sign(\langle w^*, v \rangle + b)$. When linear separability is impossible, data are projected in a feature space with greater dimension. Such transform to higher dimension Hilbert space H, can be linear or nonlinear. The kernel function K defined by (15) [37] is based on the mapping function Φ. There is no need to explicit the mapping function Φ. The radial basis function (RBF), presented in (16) is efficient for nonliear classification problem.

$$K(x_i, x_j) = \Phi(x_i). \Phi(x_j) \qquad (15)$$

$$\Phi(x, y) = exp(-g \|x - y\|^2) \qquad (16)$$

3. RESULTS AND DISCUSSION

This part aims to show the proposed method contribution in multispectral image feature extraction. The non linear separation is performed for simulated and real observations. The results of the proposed separation approach is compared to linear approaches. The proposed feature extraction method is evaluated for supervised classification application. Then, the PSS learning data set is presented and classified for different pattern transform. The classification part is performed to high resolution visible and infrared (HRVIR) observations. We use learning approach for pattern classification. The selected zone, located in north of Tunisia, presents an active scene that changes continuously due to its particular geographic situation. The image in Figure 2(a) shows various land coverts including scatter vegetation, lake, wetland, cropland, lake border and mountains. The scene is a SPOT-4 image that presents four channels. The spatial resolution is $20 \times 20 \, m$. Images are 256 gray levels. Observation images are spatially correlated. The channels correlation presented in Figures 2(b), and 2(c) shows that major data samples are in the diagonal. Corelated data will misfit the classification separblity model and impact the land cover recognition processes.
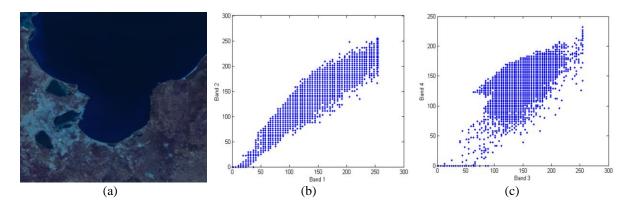

| (a) | (b) | (c) |

Figure 2. This figure are; (a) study area, (b) bands 1 and 2 correlation, (c) bands 3 and 4 correlation

3.1. Source separation for simulated mixture

In this section, we aim to study the proposed separation method on simulated mixing models. The initial sources and the joint distribution are presented in Figure 3(a) (see in Appendix). Observations are

produced by non linear mixing where the observations X_1 and X_2 are mapped to the sources by the (17). The observations are presented in Figure 3(b) (see in Appendix). Separated sources given by JADE, SOBI and the proposed separation algorithms respectively in Figure 3(c), Figure 3(d) and Figure 3(e) (see in Appendix). The JADE sources correlations to initial sources are 0.77 and 0.99, the mutual sources correlation is 0.69. For SOBI separation, the sources correlation to initial sources is 0.79 and 0.76, the mutual sources correlation is 0.13. For non linear separation sources, the correlation to initial sources are 0.99 and 0.98 and the mutual sources correlation is 0. Non linear sources decorrelation enhances the classification accuracy. Regarding the scatter plots, the estimated sources by the nonlinear approach matches the initial sources more than linear approaches as shown in Figure 3(c) (see in Appendix).

$$X_1=(1/10) * (S_1+\cos(S_2)), x2= (3/10) * (S_2+tanh(3*S_1)) \tag{14}$$

3.2. Source separation method for real observations

The observations presented in Figure 4 are correlated; the correlation is 0.9 for Band 1 and Band 2 and is 0.87 for Band 3 and Band 4. The approximated sources are presented in Figure 5. The obtained correlation is 0.03. Therefore, source separation has the advantages of ameliorating the space representation by lowering the correlation awhich will enhance the feature extraction process. Visually, we notify distinguishable classes in the sources which represent contrsted regions such as the lake (in source 1), the urban zone (in source 2), and the wetkland (in source 3). The cost function presented in Figure 6 decreasess during the sparation process and became constatnt at after 170 iterations.
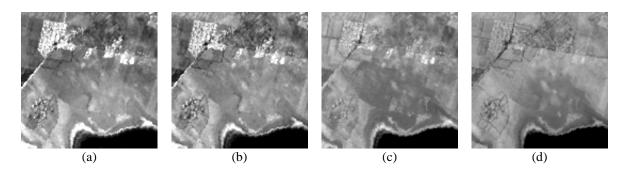

(a)   (b)   (c)   (d)

Figure 4. Multispectral image; (a), (b), (c) and (d) are SPOT4 channels

The classification results are obtained by the minimum distance algorithm. Figure 7 presented the classification results for the land categories lake (in blue), Agricultural area (in green), lake (in blue), scattered vegetation (in yellow) and Wetkland (in cyan). The band classification illustrated in Figure 7(a) provides an accuracy equal to 69%. The sources classification illustrated in Figure 7(b) provides gives an accuracy equal to 67%. For the dimension reduction method, the best source combination is the {S2, S3}. The best accuracy for this combination is 82%. Using all sources for the classification has ameliorated the classification result compared to the band classification. Moreover, using only PSS ameliorated the classification precision and avoids many manifolds and misclassified pixels. Major ground cover classes such us wetland, lake, mountains and parcels are well identified in PSS space as presented in Figure 8.
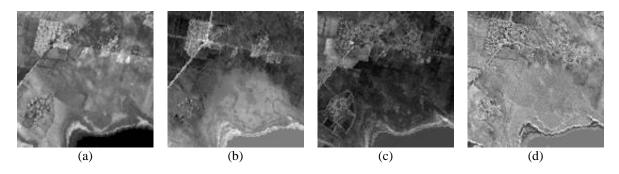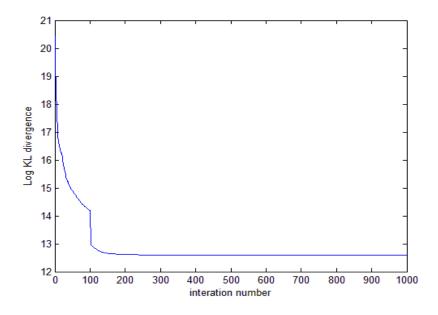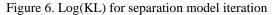

(a)   (b)   (c)   (d)

Figure 5. Source images; (a) source image 1, (b) source image 2, (c) source image 3, (d) source image 4
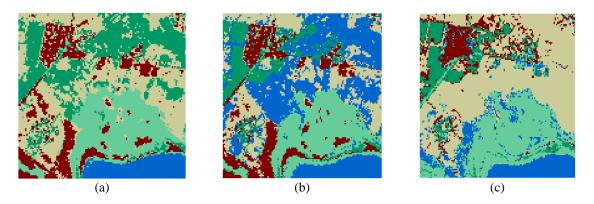
Figure 6. Log(KL) for separation model iteration

| (a) | (b) | (c) |

Figure 7. Classification results using; (a) obseravtions, (b) sources, (c) PSS

Figure 8. *SPP* and bands classification accuracy

3.3. Robustness to noise experiments

In this section, we aom to evaluate the proposed source separation method to two existing separation algorithms: SOBI and JADE. For this experimentation, white Gaussian noise contaminates the observations for different noise levels. The graph in Figure 9 demonstrates the evolution of correlation coefficients over recovered sources from noised observation. For every noise level varying between 0 and 30 db, the obtained sources are well correlated to initial sources then SOBI and JADE sources. Therefore, the proposed approach is robust in case of noisy observations. Ones we have validated the source separation and the source reduction methods, we will demonstrate their contribution in ameliorating the learning classification in the next paragraph.
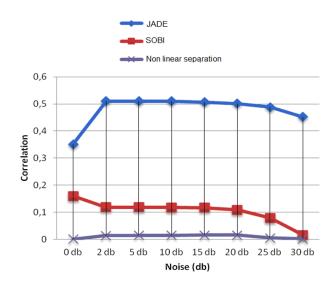

Figure 9. Correlation coefficient over source separation approaches

3.4. Learning data set

We generate, in this part, learning patches presenting the main land cover categories for the study area. The images are shown in Figure 10. Patches are 32x32 pixels. We use the PSS for the feature extraction.
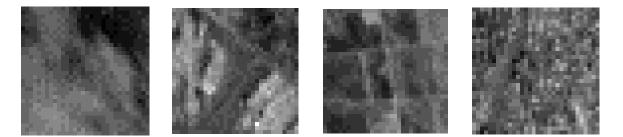

Figure 10. Learning data set samples lake wetland agriculture urban zone

3.5. Classification results

The learning set contains 405 patches and the test set contain 105 patches. Features are scaled to [0, 1] to optimize the classification results. SVM kernel used in this experimentation is the RBF that maps nonlinearly samples into higher dimension space and can handle by consequence the case when labels are mapped in a nonlinear way [40]. The classifier is parametrized by ζ and g that represent respectively the error term and the kernel parameter. Best parameters values, 4 and $0,125$ for respectively ζ and g, are determined by cross-validation. The good identification rate is 81.67%. The band classification accuracy is 75.87%. Therefore, *SPP* features provide efficient presentation for original data and allow reliable classification rate. Moreover, the *SPP* space has fewer dimension which optimize the processing time and the error impact. In the next paragraph, we will discuss the feature vectors structure using experimentations results.

3.6. Pattern transform impact on feature extraction

To show the feature selection impact on classification accuracy, we will classify the *SPP* with Gabor features and Haar features in various ways. Figure 11 details respectively the classification rate by class and the total identification rate for each feature extraction case. For the total classification rate, using Haar features gives an accuracy equal to 38.09%. Using jointly Haar and Gabor features performs 81.67% of pattern identification. Using only Gabor features gives the best accuracy result equal to 86.67%. Compared to using both Gabor and Haar features versus solely Gabor features in the *SPP* space, the general classification accuracy is better in the second case. Lake and wetland are better recognized by Gabor and Haar descriptors, while mountains, parcels and urban zone are better identified with only Gabor features. These outcomes will enhance the ckassification results for instance, for study zone presenting large presence of water like lake and wetland, we use the Gabor and Haar features; for land with little presence of water, we use only Gabor features. Existant works for land cover and land use classification reaches an overall acuuracy higher than 75% [1]-[4] using handcrafted features. Using deep semantic feature extraction reaches an accuracy higher than 90% [5]-[9]. The advatanges of our proposed methofdology, in addition to provide accurate land cover classification, is providing non correlared reduced presentation. The new reduced feature space could replace the original data in further applications including segmentation, objet detection, mutlti-resolution representation and muti-date data analysis. These applications are specific for the land cover and land use data processing and have emerging interest currently.
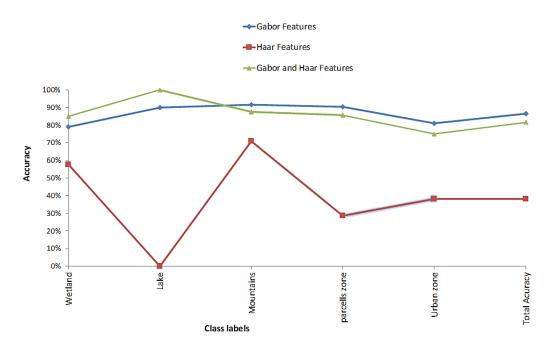

Figure 11. *SPP* classification rate for different features

4.   CONCLUSION

The multispectral land cover classification process constitutes the major goal of this work. However, the land cover variability is one of most important factors which affect the remote sensing quality data and makes classification task challenging and needing special modeling analysis methods. In this context, we presented in this paper, a new feature extraction method based on non linear separation and a new dimension reduction approach. The performance of proposed approach was demonstrated by discriminative capabilities and data quality in the classification experimental results. The outcomes of this work are firstly, the adoption nonlinear source transform approximation through neural networks to model the real and complex mixing phenomenon of multispectral data providing decorrelation data. Secondly, the classification scheme based on PSS is generated by pattern transform using Gabor and Haar technique applying to remote sensing PSS data set to identify main class labels. Using different descriptors for texture identification, the proposed approach ensure efficient feature extraction and giving most perform classification accuracy comparing with classic classifiers. To profit from the dimension reduction, the proposed approach could be used on hyperspectral data which can provides very interesting results.
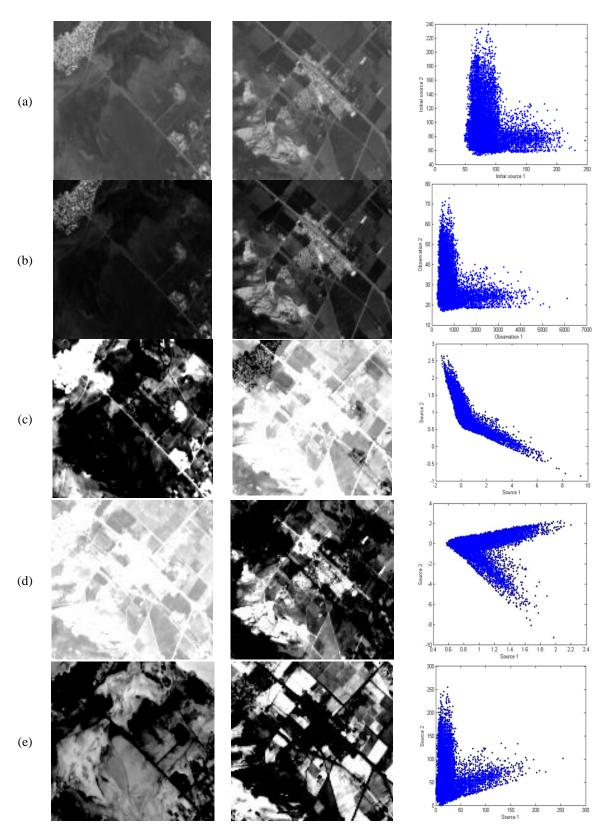
APPENDIX

Figure 3. Initial sources and approximated sources; (a) initial sources and correlation,
(b) observations and correlation, (c) JADE sources and sources correlation, (d) SOBI sources and sources
correlation, (e) non linear separation sources and correlation

ACKNOWLEDGEMENTS

REFERENCES

[1] M. Herold, X. Liu and K. C. Clarke, "Spatial metrics and image texture for mapping urban land use," *Photogrammetry and Remote Sensing*, vol. 69, no. 9, pp. 991-1001, 2003, doi: 10.14358/PERS.69.9.991.

[2] S. S. Wu, X. Qiu, E. L. Usery and L. Wang, "Using geometrical, textural, and contextual information of land parcels for classification of detailed urban land use," *Annals of the Association of American Geographers*, vol. 99, no. 1, pp. 76-98, 2009, doi: 10.1080/00045600802459028.

[3] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," *GIScience & Remote Sensing*, vol. 57, no. 1, pp. 1-20, 2020, doi: 10.1080/15481603.2019.1650447.

[4] D. Ming, J. Li, J. Wang and M. Zhang, "Scale parameter selection by spatial statistics for GeOBIA: using mean-shift based multi-scale segmentation as an example," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 106, pp. 28-41, pp. 28-41, 2015, doi: 10.1016/j.isprsjprs.2015.04.010.

[5] M. Weigand, J. Staab, M. Wurm and H. Taubenböck, "Spatial and semantic effects of LUCAS samples on fully automated land use/land cover classification in high-resolution Sentinel-2 data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 88, 2020, Art. no. 102065, doi: 10.1016/j.jag.2020.102065.

[6] M. Li, L. Wang, J. Wang, X. Li and J. She, "Comparison of land use classification based on convolutional neural network," *Journal of Applied Remote Sensing*, vol. 14, no. 1, 2020, Art. no. 016501, doi: 10.1117/1.JRS.14.016501.

[7] Y. Yu, F. Fang, Y. Liu, S. Li and Z. Luo, "Urban Land Use Classification Using Street View Images Based on Deep Transfer Network," *Urban Intelligence and Applications*, 2020, pp. 83-95, doi: 10.1007/978-3-030-45099-1_7.

[8] V. K. Rana and T. M. V. Suryanarayana, "Performance evaluation of MLE, RF and SVM classification algorithms for watershed scale land use/land cover mapping using sentinel 2 bands," *Remote Sensing Applications: Society and Environment*, vol. 19, 2020, Art. no. 100351, doi: 10.1016/j.rsase.2020.100351.

[9] L. Ding, H. Li, C. Hu and W. Zhang, "Alexnet Feature Extraction And Multi-Kernel Learning For Object-Oriented Classification Classification," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 277-281, 2018, doi: 10.5194/isprs-archives-XLII-3-277-2018.

[10] T. Siriborvornratanakul, "An Automatic Road Distress Visual Inspection System Using an Onboard In-Car Camera," *Advances in Multimedia*, vol. 2018, pp. 1-10, 2018, doi: 10.1155/2018/2561953.

[11] C. Kerdvibulvech, "Human Hand Motion Recognition Using an Extended Particle Filter," *International Conference on Articulated Motion and Deformable, Objects Articulated Motion and Deformable Objects-AMDO*, 2014, vol. 8563, pp. 71-80, doi: 10.1007/978-3-319-08849-5_8N.

[12] N. Dobigeon, A. O. Hero and J. Tourneret, "Hierarchical Bayesian Sparse Image Reconstruction With Application to MRFM," *IEEE Transactions on Image Processing*, vol. 18, no. 9, pp. 2059-2070, 2009, doi: 10.1109/TIP.2009.2024067.

[13] A. M. Djafari, "Bayesian Source Separation: Beyond PCA and ICA," *ESANN'2006 proceedings, European Symposium on Artificial Neural Networks*, Bruges, Belgium, 2006, pp. 313-322.

[14] P. Comon, "Independant component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994, doi: 10.1016/0165-1684(94)90029-9.

[15] M. A. Loghmari, M. S. Naceur and M. R. Boussema, "A Spectral and Spatial Source Separation of Multispectral Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 12, pp. 3659-3673, 2006, doi: 10.1109/TGRS.2006.882261.

[16] A. Honkela, H. Valpola, A. Ilin and J. Karhunen, "Blind Separation of Nonlinear Mixtures by Variational Bayesian Learning," *Digital Signal Processing*, vol. 17, no. 5, pp. 914-934, 2007, doi: 10.1016/j.dsp.2007.02.009.

[17] G. Burel, "Blind separation of sources: A nonlinear neural algorithm," *Neural Networks*, vol. 5, no. 6, pp. 937-947, 1992, doi: 10.1016/S0893-6080(05)80090-5.

[18] A. Honkela, T. Raiko, M. Kuusela, M. Tornio and J. Karhunen, "Approximate Riemannian Conjugate Gradient Learning for Fixed-Form Variational Bayes," *Journal of Machine Learning Research*, vol. 11, no. 106, pp. 3235-3268, 2010.

[19] H. Lappalainen and A. Honkela, "Bayesian nonlinear independent component analysis by multi-layer perceptrons," *Advances in Independent Component Analysis*, pp. 93-121, 2000.

[20] A. Honkela, "Approximating nonlinear transformations of probability distributions for nonlinear independent component analysis," *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, Budapest, Hungary, vol. 3, 2004, pp. 2169-2174, doi: 10.1109/IJCNN.2004.1380955.

[21] A. Honkela, "Advances in variational Bayesian Nonlinear Blind Source Separation," Dissertation for the degree of Doctor of Science in Technology, Helsinki University of Technology, pp. 44-61, 2005.

[22] K. L. Priddy and P. E. Kelle, "Artificial neural network: an introduction," *The international society for optical engineering SPIE Press*, Bellingham, Wachington, 2005, pp. 116-117.

[23] R. Dianat and S. Kasaei, "Dimension Reduction of Optical Remote Sensing Images via Minimum Change Rate Deviation Method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 1, pp. 198-206, 2010, doi: 10.1109/TGRS.2009.2024306.

[24] Y. Su, R. Lin and C.-C. J. Kuo, "Tree-structured multi-stage principal component analysis (TMPCA): Theory and applications," *Expert Systems with Applications*, vol. 118, pp. 355-364, 2019, doi: 10.1016/j.eswa.2018.10.020.

[25] Y. Su, Y. Huang and C. -J. Kuo, "Efficient Text Classification Using Tree-structured Multi-linear Principal Component Analysis," *2018 24th International Conference on Pattern Recognition (ICPR)*, Beijing, China, 2018, pp. 585-590, doi: 10.1109/ICPR.2018.8545832.

[26] Y. Su and C.-C. J. Kuo, "On extended long short-term memory and dependent bidirectional recurrent neural network," *Neurocomputin*g, vol. 356, pp. 151-161, 2019, doi: 10.1016/j.neucom.2019.04.044.

[27] Y. Su, K. Fan, N. Bach, C.-C. J. Kuo and F. Huang, "Unsupervised Multi-Modal Neural Machine Translation," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 10474-10483, doi: 10.1109/CVPR.2019.01073.

[28] L. O. Jimenez and D. A. Landgrebe, "Hyperspectral data analysis and supervised feature reduction via projection pursuit," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2653-2667, 1999, doi: 10.1109/36.803413.

[29] S. Kaewpijit, J. Le Moigne and T. El-Ghazawi, "Automatic reduction of hyperspectral imagery using wavelet spectral analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 4, pp. 863-871, 2003, doi: 10.1109/TGRS.2003.810712.

[30] L. H. Clemmensen, M. E. Hansen and B. K. Ersbøll, "A comparison of dimension reduction methods with application to multi-spectral images of sand used in concrete," *Machine Vision and Applications*, vol. 21, pp. 959-968, 2009, doi: 10.1007/s00138-009-0193-z.

[31] I. T. Jolliffe, "Principal Component Analysis," *New York: Springer-Verlag*, 2002.

[32] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323-2326, 2000, doi: 10.1126/science.290.5500.2323.

[33] C. Abdallah, "Application of remote sensing and geographical information system for the study of mass movements in Lebanon," Tectonics, thesis of the Université Pierre et Marie Curie - Paris VI, pp. 2007.

[34] I. Benhadj, "Observation spatiale de l'irrigation d'agrosystèmes semi-arides et Gestion durable de la ressource en eau en plaine de Marrakech," Thesis of the Doctorat de l'université de Toulouse, 'Université Toulouse III - Paul Sabatier, October 2008.

[35] G. Simonea, F. C. Morabitoa, R. Polikarb, P. Ramuhallib, L. Udpab and S. Udpab, "Feature extraction techniques for ultrasonic signal classification," *International Journal of Applied Electromagnetics and Mechanics*, vol. 15, no. 1, pp. 291-294, 2002, doi: 10.3233/JAE-2002-462.

[36] J. Zhang, Y. Wang, Z. Zhang, and C. Xia, "Comparison of wavelet, Gabor and curvelet transform for face recognition," *Optica Applicata*, vol. 41, no. 1, pp. 183-193, 2011.

[37] A. Rajaei and L. Rangarajan, "Wavelet Features Extraction for Medical Image Classification," *International Journal of Engineering Sciences*, vol. 4, pp. 131-141, 2011.

[38] E. J. Candes and D. L. Donoho, "Curvelets A Surprisingly Effective Nonadaptive Representation For Objects with Edges," *Vanderbilt University Press*, Nashville, TN, 2000.

[39] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988-999, 1999, doi: 10.1109/72.788640.

[40] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1-27, 2011, doi: 10.1145/1961189.1961199.