

TWO LEVELS FUSION DECISION FOR MULTISPECTRAL IMAGE PATTERN RECOGNITION

H. Elmannai^{a,c,*}, M. A. Loghmari^{b,c}, M. S. Naceur^{b,c}

^aEcole Supérieure des Communications de Tunis, Carthage University, 2083, Ariana, Tunisia- hela.elmannai@gmail.com

^bEcole Nationale d'Ingénieurs de Tunis, Université de Tunis El Manar (UTM), 1002, Tunisia

*Laboratoire de Télédétection et Système d'Informations à Référence Spatiale

KEY WORDS: Source separation, Support Vector Machine, Feature extraction, Decisional fusion, Pattern recognition

ABSTRACT:

Major goal of multispectral data analysis is land cover classification and related applications. The dimension drawback leads to a small ratio of the remote sensing training data compared to the number of features. Therefore robust methods should be associated to overcome the dimensionality curse. The presented work proposed a pattern recognition approach. Source separation, feature extraction and decisional fusion are the main stages to establish an automatic pattern recognizer.

The first stage is pre-processing and is based on non linear source separation. The mixing process is considered non linear with gaussian distributions. The second stage performs feature extraction for Gabor, Wavelet and Curvelet transform. Feature information presentation provides an efficient information description for machine vision projects.

The third stage is a decisional fusion performed in two steps. The first step assign the best feature to each source/pattern using the accuracy matrix obtained from the learning data set. The second step is a source majority vote. Classification is performed by Support Vector Machine. Experimentation results show that the proposed fusion method enhances the classification accuracy and provide powerful tool for pattern recognition.

1. INTRODUCTION

The major goal leading researches in multichannel data is to find a suitable presentation for data analysis, pattern recognition and related applications. In this framework, many interesting works within the source separation process and pattern classification axes have arisen (Dobigeon and al., 2009) (Halimi and al., 2011). The first axe, is the mixing model relating observations to land cover. The phenomenon is non linear due to the radiance emitted by heterogeneous soil and by reflections over atmosphere layers and clouds (Loghmari and al., 2006). Many existent source separation models are based on various assumptions and simplifications. Non linear model is the most realistic case. Therefore, non linear source separation pre-processing has relevant importance to establish a powerful pattern recognition tool by avoiding channel correlation for feature extraction (Yaltmann, 2013).

The second axe is pattern recognition and land identification. Pattern recognition focuses in finding reliable feature description for land cover. For this purpose, various morphological, mathematic transform and contour descriptors exist (Abdallah, 2007). Main researches aim to reach high classification performance for satellite scene identification which is necessary in case of fire, inundation and fast land change detection. Associating many descriptors could improve the pattern recognition accuracy.

This paper starts by presenting the methodology algorithm in section 2. The section 3 presents briefly the non linear separation method and assumptions. Section 4 presents the feature extraction method. Section 5 details the decisional fusion process. In section 6 we explore the experimentations

results on multispectral scene. Section 7 concludes and presents future research directions.

2. METHODOLOGY PRESENTATION

Pattern recognition refers to identify land cover semantic signification. Traditional methods fail due to the dimensionality curse. To have an efficient recognition tool for land pattern, we need the contribution of many computational intelligences and algorithms. The presented work aims to provide reliable pattern recognition method for remote sensing image. The tool is based on learning strategy which avoid land identification manifolds. The proposed method follows three steps. These steps are pre-processing, feature extraction and decisional fusion. The pre-processing provides uncorrelated new-channels image. Feature extraction is performed for the learning source set. The process of feature extraction includes various descriptors for enhanced pattern recognition. The decisional fusion is based on Support Vector Machine (SVM) classification and has two level: feature level fusion and source fusion level. Let consider the following sets and notations:

The feature set: $D = \{d_j\}$

The label set: $L = \{L_k\} = \{Label_1, Label_2, Label_3, \dots\}$

The source set: $E = \{S_i\}$

$S_{i,j}$: Pattern decision for the source S_j given the feature j .

The general approach is detailed in Figure 1. The fusion stage is based on three algorithms: SVM classification, feature fusion level and two levels source fusion level.

* Corresponding author

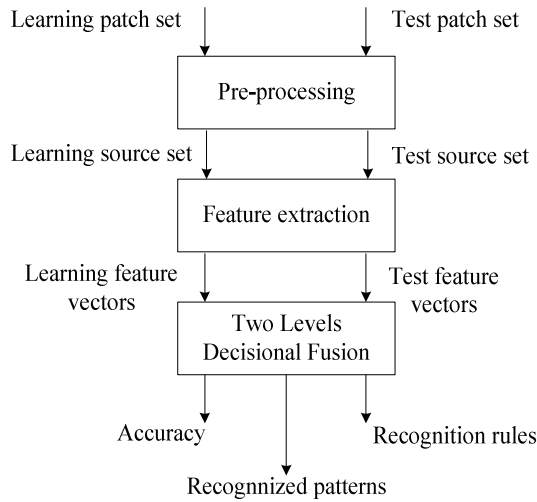


Figure 1. General approach presentation

3. NONLINEAR BAYESIAN SOURCE SEPARATION

Physical modeling techniques have been used for imaging phenomenon analysis. Source separation aims to recover hidden sources from measured signals. Linear mixture models consider a linear mixture of spectral signatures belonging to the present materials within the pixel. Some other linear methods are based on assumptions made on source distributions. The Principal Component Analysis (PCA) and Independent Component Analysis (ICA) models are linear and are based on sources independence assumption (Djafari, 2006) (Honkela and al. 2007). More realistic models consider non linear mixture due to distortions within sensors and atmosphere. Based on non linear counterpart of the PCA algorithm (Honkela and al. 2007), we propose a non linear mixture model including noise term. The latent factors are considered independent and have Gaussian distributions. The nonlinearity is performed by two-layer perceptron. Similar implementations used gradient descent updating rule and mutual information minimizing (Burel, 1992). Other implementations are based on back-propagation algorithm and the entropy as a measure of dependency (Honkela and al., 2010).

The presented approach uses Bayesian inferences for estimating the unknown parameters from their priors and uses conjugate gradient. Let consider X the N observed signals $X(t)=[x_1(t), x_2(t), \dots, x_N(t)]^T$. The latent sources are $S(t)=[s_1(t), s_2(t), \dots, s_M(t)]^T$, T is the transposition operator. The nonlinear mapping is presented by the Equation 1.

$$f(S(t)) = B \tanh(AS(t) + a) + b + n(t) \quad (1)$$

$f: \mathcal{R}^M \rightarrow \mathcal{R}^N$ is a non linear function, H denotes the neurons in the hidden layer, $A \in \mathcal{R}^{H \times M}$ and $B \in \mathcal{R}^{N \times H}$ are weight matrices. The hyperbolic tangent \tanh is the activation function for the hidden layer. $a \in \mathcal{R}^H$ and $b \in \mathcal{R}^N$ denote the bias vectors for hidden and output neurons respectively. n is a Gaussian noise. The Bayesian Ensemble learning is performed by numerically minimizing a cost function. Considering the relative entropy between the posterior probability density functions and their approximations. Posterior PDF are updated by minimizing the Kullback-Leibler (KL) divergence between the approximated

PDF $q(\theta | X)$ and the posterior PDF $p(\theta | X)$. KL divergence is defined by Equation 2. θ denotes the unknown parameters (source, noise, MLP weights and biases).

$$KL(q(S, \theta) \| p(S, \theta | X)) = \int q(S, \theta) \log\left(\frac{q(S, \theta)}{p(S, \theta | X)}\right) d\theta \quad (2)$$

4. FEATURE EXTRACTION

Feature extraction allows a consistent presentation for information through morphological, mathematical transform and statistics. Pattern descriptors need efficient and well chosen descriptors to improve the pattern identification rate. Content Based Image Retrieval (CBIR) techniques (Datta and al., 2008) are based on retrieving visual similarity through visual characteristics. Similarity measures the textural, shapes or colour information's resemblance. The measure may be fuzzy or deterministic and may use supervised or mi-supervised learning. Associating many features has proven to ameliorate the image retrieval accuracy.

In this context, we aim to establish a framework for land cover pattern recognition. Salient patterns are agricultural parcels urban area, mountains, wetlands and lakes. For this purpose, we have to choose the most suitable feature in a limited set of considered patterns. The feature set will contain Wavelets, Curvelets and Gabor features.

4.1 Wavelet features

Wavelet features are extracted by decomposing the image into approximate, horizontal, vertical and diagonal coefficients. The decomposition is parameterized by a decomposition level. This transform acts as a multi-scale differentiator presenting image singularities (Zhang and al., 2011) in many orientations and scales. It allows low-frequency and high-frequency image presentations. Low frequency sub-band gives an image approximation, other sub-bands give a high or low frequency information in horizontal or vertical directions (Rajaei and al., 2011).

4.2 Curvelets features

For edge image singularities, wavelets have proven to be inefficient. Candes and Donoho have introduced the curvelet transform in (Candes and al., 2000). Discrete curvelets transform present edges and other singularities along curves much more efficiently. Discrete Curvelet Transform was introduced by Candes and Donoho in (Candes and al., 2000). For the transform implementation, Candes and al have proposed (Candes and al., 2006) two forms based respectively on unequally spaced fast Fourier transform and wrapping based fast curvelet transform. The second transform is more robust and fast. The frequency plane is divided into annular concentric rings. Each ring is portioned into annular edges. Edges correspond to scale levels and allow multiscale and multiorientation image analysis. Directional resolution is doubled at each scale.

4.3 Gabor features

Texture pattern is relevant in land pattern identification. Many regions have singular visual characteristic such as urban areas and mountains that distinguish it from other regions. The multichannel filtering based on Gabor filters have been widely used for texture identification. They act as a multichannel filters

offering different scale frequencies and directions. The usual used values are five scales and eight orientations.

5. TWO LEVELS FUSION DECISION

In this work, the feature vector selection test will contain Gabor, Haar transform and curvelet transform features (Zhang and al., 2011). Considering the feature set $D = \{d_j\} = \{d_{Wavelet}, d_{Curvelet}, d_{Gabor}\}$, the source set $S = \{S_i\}$, the decision set $L = \{Label_k\} = \{Label_1, Label_2, Label_3, \dots\}$ presenting the pattern labels possibilities. Each feature is obtained through the correspondent transform and summarize the transform statistics in term of mean, variance and standard deviance. For each patch P , $S_j(P)$ is the pattern recognition function for the source S_i based on the feature category j .

Fusion levels may concern data, features or decisions. The first fusion level concerns data and may include multi-dates data, panchromatic, and other informations. Feature fusion level is based on data characteristics. Data decision fusion level concerns high level data representation such as classification, feature detection or pattern recognition. Actual methods used data level or feature level. In the presented work we present two levels decision fusion method. For this purpose, the SVM is used for the learning classification task. SVM provide the accuracy matrix that will be used for the feature decision level. Then the upper decision will concerns sources. The two levels fusion framework is detailed in Figure 2. The SVM principle, feature decision level and source decision level will be detailed in next parts.

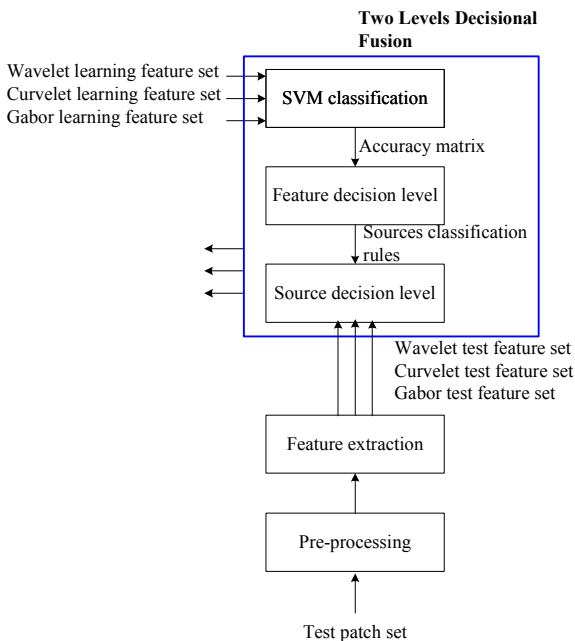


Figure 2. Two levels decisional fusion.

5.1 Support Vector Machine classification

Learning machines outperform traditional method in pattern classification. SVM have proven their effectiveness for remote sensing image classification (Kovacevic and al., 2009). SVM were introduced by Cortes and Vapnik (Vapnik, 1999). The classification is based on generating a decision function defined by a vectors subset from the learning database. An hyperplane separability, in binary classification, determines the vector label.

In case of nonlinear separability, the classification process projects the training data into higher dimensional space and then specifies a maximum-margin separating hyperplane in the projected space (Zammit and al., 2007). The mapping function is denoted by $F: \mathcal{R}^d \rightarrow H$, where H is a higher dimension Hilbert space. There is no need to explicit the mapping function Φ . Let consider the kernel function K defined by Equation 3.

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (3)$$

The decision function is given by Equation 4, sign is denoted Sgn .

$$\sigma(x) = sign \left(\sum_{i \in \{1, \dots, k\}} y_i \varepsilon_i K(x, x_i) + b \right) \quad (4)$$

Commonly used kernel function are linear, polynomial, and Radial Basis Function (RBF). RBF kernel can handle the case when the relation between class labels and attributes is non linear and has fewer numerical difficulties (Mitra and al., 2004).

5.2 Feature fusion level

For each source, the SVM classifier provides an accuracy matrix defined by the accuracy given by using the feature set elements to recognize the pattern set E . The accuracy matrix for the source i denoted $M_i = \{\alpha_{kj}\}$, k is the pattern name and j is the feature name.

This fusion scheme allows to notify relative importance for each feature category for each pattern recognition. The decisional fusion is based on the following principle: for each pattern, the higher accuracy determines the most credible feature for the correspondent pattern. α_{kj^*} denotes the higher accuracy for the pattern k which is given by the feature j^* . The α_{kj^*} should be sorted from maximum to the minimum to produce a rule tree.

We deduce another empty matrix noted $M_i^* = \{\beta_{kj}\}$ by putting $\beta_{kj^*} = 1$ for $j=j^*$, and $\beta_{kj} = 0$ for $j \neq j^*$. Then the classification rule expression for a source patch $S_i(P)$ is $S_i(P) = j^*$ if $\beta_{kj^*} = 1$. The lower rule is given by the next α_{kj} in the sorted list.

5.3 Source fusion levels

The source fusion is performed by the majority vote principle. Let consider a patch P and M sources. Reminding that $S_i(P) = j$ is the source decision for the patch P given by the previous decision step. The indicator function X_{ij} is defined by:

$$X_{ij}(P) = 1 \text{ if } S_i(P) = j \text{ and } X_{ij}(P) = 0 \text{ if } S_i(P) \neq j.$$

The majority vote principle computes for each label k the next expression Equation 5.

$$X_k^S(P) = \sum_{i=1..M} (X_{ki}^i(P)) \quad (5)$$

The majority vote decision is therefore:

$$S(P) = k \text{ if } M_k^S(P) > M/2 \quad (6)$$

Non decisional cases are resolved for the accuracy matrix M_i by taking the feature that belongs to the next greater accuracy.

6. EXPERIMENTS AND ANALYSIS

Initial observations are High-Resolution Visible and Infra-Red (HRVIR) image. Figure 3 presents SPOT4 composite images for the zone of interest. The multispectral image are 4 channels, the spatial resolution is 20mx20m. The pattern identification process work is used for macro pattern identification (urban areas, parcels, moutains, ...). The selected geographic zone is in

north-est of Tunisia and is reputed of its heterogeneity and various patterns.

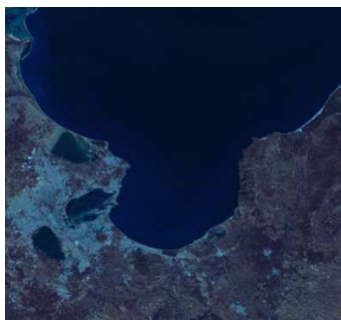


Figure 3. Study area

Two experiments are performed and evaluated: non linear source separation method and pattern recognition. For separation experimentation, supervised classification performance and robustness to noise are detailed and compared to other known separation methods.

6.1 Source separation experimentation

The first experimentation part aims to evaluate the source separation process which is an important step in the proposed methodology. The Minimum Distance classification is used as a classifier test example for supervised classification. We tend to analyse the separation contribution impact in land cover identification.

Remembering that the source separation process is based on Bayesian inferences and the Gaussian assumption for sources and latent factors. Considering the Expectation-Maximization algorithm (Dempster and al., 1977), the Gaussian models for observation show mixed distributions in Figure 4. The source separation result shows more identifiable distributions (Figure 4).

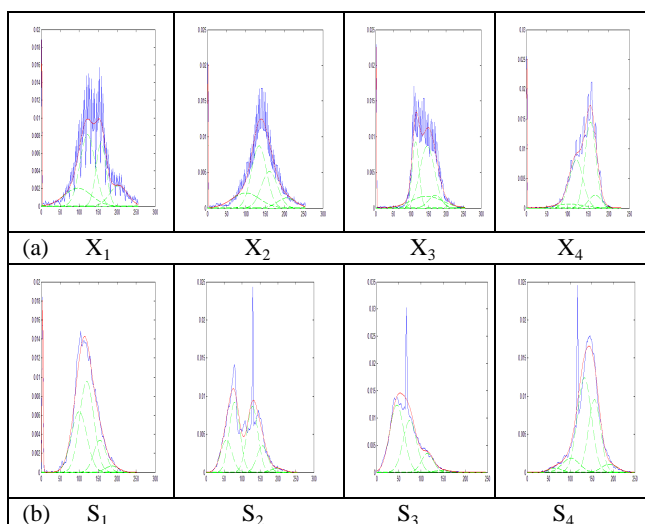


Figure 4. Data distributions before (a) and after non linear source separation(b)

The test zone for the Minimum Distance test is illustrated in Figure 5. We use the training sites presented in Table 1.

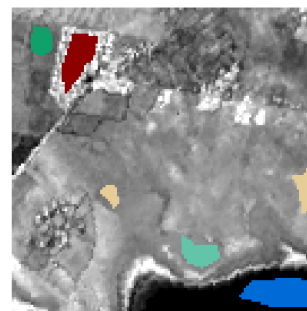


Figure 5. Training sites for supervised classification

Color	label	Pixel number
Blue	Lake zone	2073
Green	Agricultural zone	545
Brown	Urban area	1472
Cyan	Wetland	607
Yellow	Scattered vegetation	1407

Table 1. Minimum Distance samples number

The supervised classification using the Minimum Distance gives the classified images presented in Figure 6.

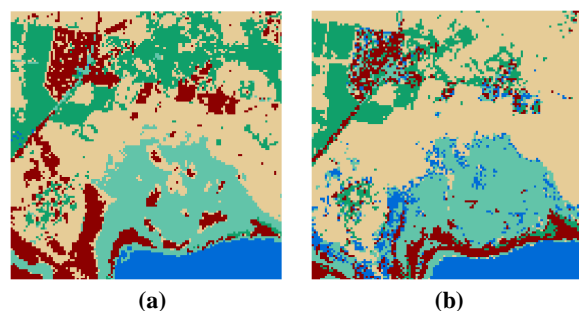


Figure 6. Image clustering. (a) Band classification. (b) Source classification.

Thanks to the given ground-truth, the Minimum Distance classification shows that source classification provide a good identification rate of about 81% while band classification is 64%. In fact, non linear source separation provide uncorrelated data which is less than 0.03. The observation correlation is more than 0.65.

In the following, we compare the proposed separation approach with two standard source separation techniques: Second Order Blind Identification (SOBI) and Joint Approximate Diagonalization of Eigen matrices (JADE). These algorithms are linear approaches and use second and higher order statistics. The experimentation consists on contaminating observation by a white Gaussian noise and evaluate the correlation coefficients for estimated sources and the original ones. The noise power in decibels is leaded by 7 levels varying from 0 to 25 db. The normalized error over the linear approach is lower than 0.1 for all noise levels. The proposed approach is more robust than

SOBI and JADE algorithms and has higher performance in case of noised observation (Figure 7).

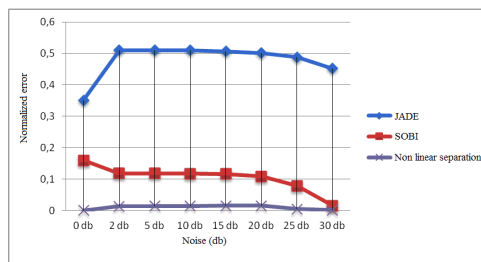


Figure 7. Normalized error over source separation approaches

The proposed source separation process enhances data reliability by avoiding information correlation and enhancing robustness to noise. These two advantages will improve the pattern recognition result which will be shown in the next experimentation part.

6.2 Pattern recognition results

The process of source separation and feature extraction is performed for learning data set. Figure 8 presents learning patches. The learning data set contains 405 image 32x32 exemplars. The considered classes are wetland, agricultural parcels, mountains, lake and urban areas.

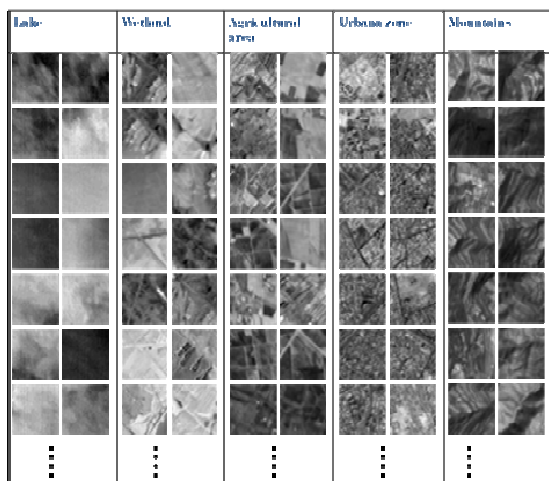


Figure 8. Learning data base samples

Reminding the pattern recognitions notations:

$$D = \{d_i\} = \{d_{\text{Haar}}, d_{\text{Curvelet}}, d_{\text{Gabor}}\},$$

$$S = \{S_1, S_2, S_3, S_4\}$$

$$L = \{\text{Label}_k\} = \{\text{Wetland}, \text{Lake}, \text{Mountains}, \text{Parcels}, \text{Urban areas}\}$$

The accuracy matrix for the source S_1 is given by Table 2. Bold rates represent the max accuracy for each row (pattern). The sorted accuracy list is:

$$\text{list} = \{ \alpha_{\text{urban Curvelet}} = 94,40\%, \alpha_{\text{Mountain Gabor}} = 93,20, \alpha_{\text{Lake Gabor}} = 91,21\%, \alpha_{\text{Parcels Gabor}} = 90,48\% \text{ and } \alpha_{\text{Wetland Gabor}} = 77,92\% \}.$$

We deduce the empty matrix noted $M^* = \{\beta_{kj}\}$ given by Table 3. The deduced identification pattern rules for S_1 and for a patch test P:

if ($M^*_i(\text{urban}, \text{curvelet})=1$) then $S_i(P)=\text{Curvelet}$
 else if ($M^*_i(\text{Mountain}, \text{Gabor})=1$) then $S_i(P)=\text{Mountain}$
 else if ($M^*_i(\text{Lake}, \text{Gabor})=1$) then $S_i(P)=\text{Lake}$
 else if ($M^*_i(\text{Parcels}, \text{Gabor})=1$) then $S_i(P)=\text{Parcels}$

else if ($M^*_i(\text{Wetland}, \text{Gabor})=1$) then $S_i(P)=\text{Wetland}$.

Else considerate the additional rules generated by the second superior rate by row α_{ki} .

The principle is to extract the values, sorted them and deduce additional rules until finding the patch pattern.

	Haar Feature	Curveletes Feature	Gabor Feature
Wetland	50,30%	76,60%	77,92%
Lake	76 %	20,55%	91,21%
Mountains	78,18%	80,33%	93,20%
Parcels	74,19%	84,11%	90,48%
Urban areas	71,03%	94,40%	79,17%

Table 2. S_1 accuracy matrix

	Haar Feature	Curveletes Feature	Gabor Feature
Wetland	0	0	1
Lake	0	0	1
Mountains	0	0	1
Parcels	0	0	1
Urban areas	0	1	0

Table 3. S_1 empty matrix

The feature fusion is applied for all sources. $S_1(P)$, $S_2(P)$, $S_3(P)$ and $S_4(P)$ are the sources decision for the patch pattern P. The majority fusion vote consider that all sources have the same reliability. And the chosen pattern belongs to the most voted one. For non decisional case the patch is considered non recognized. The given accuracy for 105 test patch is 92%.

The accuracy matrix shows the Gabor features contribution for land covers pattern identification. Haar wavelets are not efficient in our case. This leads to an adaptable classifier that depend on the a-priori information about the land covers. In fact, in case of land cover with little urban areas superficies, Gabor feature are efficient. Linear separation pre-processing allows source decorrelation and enhances therefore the decisional performance. In the other hand, each source presents many distinguishable land thematics which allows a direct relation between land label and land pattern.

Classification rules depends on the learning set and allow an intelligent pattern recognition tool. Joining all sources decision provide robustness to the pattern recognition tool. The method is general and could be ameliorated with other features category and other data sources types like multi-resolution image, multirate image and land-elevation.

7. CONCLUSION

The proposed work presents a new approach for multispectral data analysis and pattern recognition. The pre-processing stage presents a nonlinear approximation model to recover the real unseen sources. Recovered sources are not correlated and present the land cover more precisely than observations. Then feature extraction associate many feature category for an enhanced data presentation. The pattern recognition task in performed in two level: decisional fusion within the feature vectors, and then decisional fusion within sources.

The approach shows a clear improvement compared to classification of initial images. This application is of utmost importance in the multispectral image analysis. Future works

will interest micro structure identification and will use higher spatial resolution and hyperspectral images.

REFERENCES

- Abdallah, C., 2007. Application of remote sensing and geographical information system for the study of mass movements in Lebanon. thesis of the Université Pierre et Marie Curie - Paris VI, pp. 69-75.
- Altmann, Y., 2013. Nonlinear spectral unmixing of hyperspectral images. Thesis for the degree of Doctorat de l'Université de Toulouse, University of Toulouse, France.
- Burel, G. , 1992. Blind separation of sources: A nonlinear neural algorithm. *Neural Networks*, vol. 5, pp. 937-947.
- Candes, E., Demanet, L., Donoho, D. and Ying, L., 2005. Fast Discrete Curvelet Transform. *Multiscale Modeling and Simulation*, vol. 5, pp. 861-899.
- Candes, E. , and Donoho, D. L. , 2000. Curvelets A Surprisingly Effective Nonadaptive Representation For Objects with Edges, Vanderbilt University Press, Nashville.
- Datta, R., Joshi, D., Li, J. and Wang, J. Z., 2008. Image retrieval: Ideas, influences, and trends of the new age, *ACM Computer. Survey* 40(2), pp. 1-60.
- Dempster, A. P., Laird, N. M., and Rubin, D. B., 1977. Maximum likelihood from incomplete data via the em algorithm, *Journal of the Royal Statistical Society, Series B*, pp. 1-38.
- Djafari, A. M., 2006. Bayesian Source Separation: Beyond PCA and ICA. In: *European Symposium on Artificial Neural Networks proceedings*, Belgium, pp. 313-322.
- Dobigeon, N., Hero, A. O. and Tourneret, J-Y., 2009. Hierarchical Bayesian Sparse Image Reconstruction With Application to MRFM. *IEEE transactions on Image Processing*, Vol. 18, No. 9, pp. 2059-2070.
- Halimi, A., Y. Altmann, Y., Dobigeon, N. and Tourneret, J-Y. , 2011. Nonlinear Unmixing of Hyperspectral Images Using a Generalized Bilinear Model. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 49, No. 11, pp. 4153 - 4162.
- Honkela, A., Raiko, T., Kuusela, M., Tornio, M. and Karhunen, J. 2010. Approximate Riemannian Conjugate Gradient Learning for Fixed-Form Variational Bayes", *Journal of Machine Learning Research*, vol. 11, pp. 3235-3268.
- Honkela, A., Valpola, A., Ilin, A. and Karhunen, J., 2007. Blind Separation of Nonlinear Mixtures by Variational Bayesian Learning. *Digital Signal Processing Special Issue on Bayesian Source Separation*, Vol. 17, no. 5, pp. 914-934.
- Kovacevic, M., Bajat, B., Trivi, B. and Pavlovic, R., 2009. Geological Units Classification of Multispectral Images by Using Support Vector Machines. *International Conference on Intelligent Networking and Collaborative Systems proceedings*, pp. 267- 272. Barcelon, Spain.
- Loghmani, M. A., Naceur, M. S. and Boussema, M. R., 2006. A Spectral and Spatial Source Separation of Multi-spectral Images. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 44, No. 12, pp. 3659 - 3673.
- Mitra, P., Shankar, B. U. and Pal, , S. K., 2007. Segmentation of multispectral remote sensing images using active support vector machines, *Pattern Recognition Letters* 25, pp. 1067-1074.
- Rajaei, A. and Rangarajan, L., 2011. Wavelet Features Extraction for Medical Image Classification. *International Journal of Engineering Sciences*, Vol. 4 , pp. 131-141.
- Vapnik, V. N., 1999. An Overview of Statistical Learning Theory, *IEEE Transactions on neural networks*, vol. 10, No. 5, pp. 988 - 999.
- Zammit, O., Descombes, X. and Zerubia, J., 2007. Apprentissage non supervisé des SVM par un algorithme des Kmoyennes entropique pour la détection de zones brûlées, *Colloque GRETSI proceedings*, Troyes, France, pp 11-14.
- Zhang, J., Wang, Y., Zhang, Z. and Xia, C., 2011. Comparison of wavelet, Gabor and curvelet transform for face recognition. *Optica Applicata*, Vol. XLI, No. 1, pp. 183-193.