An approach using wavelet transform for land cover changes detection on remote sensing data

Sonia Bouzidi, Isabelle Herlin

AN APPROACH USING WAVELET TRANSFORM FOR LAND COVER CHANGES DETECTION ON REMOTE SENSING DATA

Sonia Bouzidi[1], Isabelle Herlin[2]

[1] *LSC FRE CNRS 2494*
Evry University 40, rue du Pelvoux
91020 Evry Cedex
`Bouzidi@iup.univ-evry.fr`

[2]*AIR Project*

INRIA-BP 105 – 78153

Le Chesnay Cedex

`Isabelle.Herlin@inria.fr`

Abstract:

The paper addresses a new application of time-frequency representation in classification of non-stationary signals. It defines an approach to characterize curves describing the temporal reflectance profiles related to land cover behavior. These curves are obtained from processing multi-resolution remote sensing data and then projected on a representation space improving discrimination between different land cover types. This approach can be applied in a scenario dedicated to the detection of land use changes caused by erosion.

Key words: time-frequency transformation, signal classification, land cover change, remote sensing

1. Introduction

Time-frequency transformation has proved its efficiency for non stationary signals analysis in different application fields such as : speech analysis[1,2], biomedical engineering[3], seismology, imagery[4], ...
One of the most used techniques to perform signal transformation is the Wavelet Transform and more particularly Discrete Wavelet Transform (DWT) which filters noise and provides a compact representation of a signal making it easier the extraction of information[5].

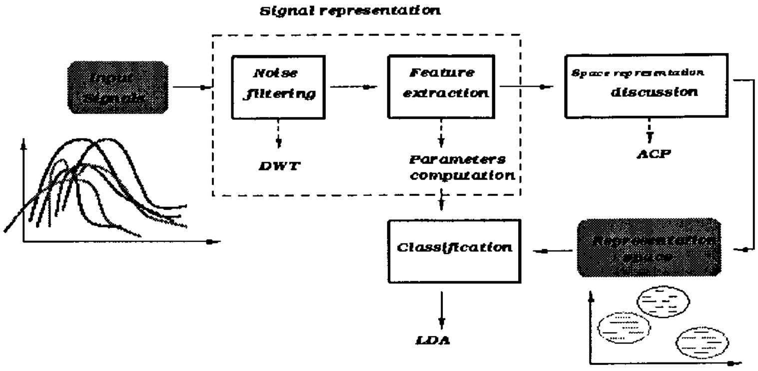
This paper presents another application area of time-frequency transformation. It addresses the use of this technique in order to characterize curves describing the temporal reflectance profiles related to land cover behavior. These temporal profiles are estimated from remote sensing data using a fusion scheme of high spatial resolution data provided by Landsat satellite and high temporal frequency images from NOAA/AVHRR. This estimation is performed locally. We then obtain a set of curves for each land cover type by varying the spatial location. We aim to project these profiles into a representation space where the discrimination between land cover classes (groups of curves related to each land cover type) is easier than in the initial time domain. The final objective is to improve land cover monitoring on large areas by characterizing each land cover type through its reflectance temporal profile shape. This characterization can be applied to detect land cover changes, caused by natural problems as erosion.

The paper is organized as follows: In section2, we describe the developed approach with the different steps: signal transformation using DWT (subsection 2.1), feature extraction (subsection 2.2), discussion of the representation space choice and the classification method (subsection 2.3). Section 3 describes the process of curves estimation from remote sensing data. In section 4, we discuss the obtained results. The section 5 is devoted to describe a scenario of application for this work to land cover changes detection. Finally we conclude and present future work in section 6.

2. Approach developed

The input data of our process are curves describing the temporal behavior of different land cover types. These curves represent reflectance evolution in function of time (day of the year) in the visible and Near-Infrared (NIR) channels of NOAA/AVHRR sensor. They are obtained by modeling the NOAA pixels composition as

explained in section 3. We aim to project these curves onto a feature space where classes of land cover will be better discriminated.
To achieve curves classification the approach illustrated in figure 1 is followed.



The first stage is the signal representation which consists on applying a Discrete Wavelet Transformation in order to denoise the signal, then a feature extraction to characterize it. The second stage is the classification of the feature vectors.

2.1 Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) provides a time-frequency representation of a signal. It transforms a time domain discritized signal *x(n)* of finite length $\{1, \ldots, 2^N\}$ into its corresponding wavelet domain. This is done through a process called ``sub-band codification" which is done through digital filter techniques. It allows the signal decomposition into an approximation and detail information. The approximation is then further decomposed using the same wavelet decomposition transform. This achieved by successive high-pass and low-pass filtering of the time domain signal and is defined by the following equations:

$$Y_{high}[k]=\sum_n x[n]g[2k-n] \qquad (1)$$

$$Y_{low}[k]=\sum_n x[n]h[2k-n] \qquad (2)$$

Where $Y_{high}[k]$, $Y_{low}[k]$ are respectively the outputs of the high-pass *(g)* and low-pass *(h)* filters, after subsamling by 2. A set of different wavelet families have been proposed in the literature. In our implementation, the four coefficients wavelet family (DAUB 4) proposed by Daubechies [6] is used.

2.2. Feature extraction

After performing the DWT, statistics over the set of wavelet coefficients are used. We compute different parameters to study the profiles features:

- the mean of the absolute values of the coefficients in each sub-band. These features provide information about the reflectance distribution;
- the standard deviation in each sub-band, this provides information about variation in the frequency distribution.

- the maximal and the minimal values and their positions in the approximation sub-bands. These features allow a time-scale localization and provide information about the curves shape.

The observation and comparison of these parameters in the different sub-bands at each decomposition level allow the choice of the more significant parameters as well as the decomposition level supposed sufficient for the signal representation without loosing information. Since each profile is described by two curves in the visible (channel1 of NOAA/AVHRR) and in the Near Infrared (channel 2 of NOAA/AVHRR), feature extraction is performed in the two channels. So that each land cover profile is characterized by the parameters previously described and computed in the two NOAA channels.

2.3. *Classification and representation space*

The data (Temporal profiles) are divided into a training set and a test set. The training set is used to define the classes and to choose the representation space. The test set is used to evaluate the classification performance.

Let N be the number of training profiles (x_i^k, c); for each land cover class c ; $c = 1, \ldots, C$ *where* C is the number of land cover types, $i = 1, \ldots, N$; $k = 1$ *and* 2 (NOAA channels). For each profile (x_i^1, x_i^2, c), we compute the feature vector as it is described in the subsection 2.2, then we perform classes definition. Two representation spaces are discussed:

- The first one consists in applying Principal Components Analysis (PCA) in order to reduce dimensionality and to produce an uncorrelated feature set. Hence, the feature vectors are projected on the principal components space where classes are defined. In this new space, a Linear Discriminant Analysis (LDA) is applied. It is a linear statistic classification method which maximizes the ratio of between class variance to the within class variance in order to obtain decision regions between the different classes of data. It is a question of finding linear transform A, so that after its application the scatter of sample vectors is minimized within each class and the scatter of mean vectors around the total mean vector is maximized simultaneously.
- The second case consists in applying LDA directly in the space of the original features (without PCA application). In our case, this allows us to test the relevance of reducing dimensionality by PCA and to evaluate its effect on the LDA performance.

2.3.1 Classification method description

The classification step is based on defining a criterion to assign a temporal profile of reflectance, described by its features vector, to a class of land cover. Each class c is defined by its mean vector .

$$X_c = \frac{1}{N}\sum_{i=1}^{N} X_{i,c} \qquad (3)$$

where $X_{i,c}$ is the original feature vector or its projection on the principal component space after a PCA application.

The within class variance is estimated by the covariance Matrix (4)

$$S_w = \sum_{c=1}^{C}\frac{1}{C}\sum_{i=1}^{N}\frac{1}{N}(X_{i,c} - X_c)(X_{i,c} - X_c)^T \quad (4)$$

The between class variance is *(5):*

$$S_b = \sum_{c=1}^{C}\frac{1}{C}(X_c - X)(X_c - X)^T \qquad (5)$$

Where

$$X = \sum_{c=1}^{C}\frac{1}{C}X_c$$

The maximization of the ratio $\frac{S_b}{S_w}$ allows the determination of the linear transformation A improving classes discrimination. A new temporal profile of reflectance (or a temporal profile from the test set) Y is assigned to the class for which $d(A^T Y; A^T X_c)$ is minimal, with d the euclidian distance.

3. Curves determination

Identifying and monitoring land cover components is depending on spatial resolution and temporal frequency of observations.

High spatial resolution data provided by Landsat TM or SPOT improved with ground truthing give a good information about land cover. However, it is too difficult and very expensive to frequently obtain this kind of data for important surfaces.

To study progressive phenomena, we can exploit NOAA/AVHRR sensors which provide daily acquisitions covering large areas. Nevertheless, due to the size of a pixel (1.1 km), which is generally larger than a single field, monitoring vegetation with NOAA/AVHRR data implies a mixture modeling for pixels values. We assume a linear model (6) where the reflectance of a pixel is considered as a sum of the individual land covers reflectances weighted by their area proportions [7]. This relation is considered in the visible and Near InfraRed channels:

$$r_i^k(t) = \sum_{j=1}^{C} \rho_{ij} R_j^{k(loc_i)}(t) \quad (k=1,2) \qquad (6)$$

where:

- C is the number of land cover types,
- $r_i^k(t)$ the reflectance of a NOAA pixel i in channel k $(k = 1 or 2)$, at date t.
- $R_j^{k(loc_i)}(t)$ the individual reflectance of land cover type j computed locally around the pixel i in channel k $(k = 1$ or $2)$, at date t.
- ρ_{ij} the proportion of the land cover j within a NOAA pixel i.

We consider an area for which ground truth and remote sensing data are simultaneously available. So that a precise land cover thematic classification is obtained for this area at high spatial resolution (of Landsat). Then, after geometrical registration, we superpose the Landsat classification image and the NOAA images [8]. Consequently, for each NOAA pixel, its composition in terms of percentage of land cover can be obtained by directly counting pixels on the Landsat classification image. We then obtain the ρ_{ij} values for this learning area.

At each date t, for each pixel i, we consider its neighbors. The neighborhood size is depending on the number of land cover types, in our case we consider 8 neighbors of a pixel i. So we obtain a linear system of 9 equations. The are observed on the NOAA sequence of images; knowing the ρ_{ij}, the inversion of the linear system allows the estimation of individual values for this date. By repeating this process for all the dates, the temporal profiles $R_j^{k(loc_i)}(t)$ are computed. So that for each land cover type j and for each NOAA channel (1 and 2), we obtain as many curves as considered NOAA pixels.

4. Experimental results and discussion

The developed approach has been applied on a test area of the Mkomazi basin in South Africa. The thematic classification obtained for this region at Landsat resolution presents four classes of land cover: water, grass, forest and bare soil. We first produce temporal profiles of reflectance at it is previously explained. The figures 2 and 3 represent an example of results obtained for grass. They show a set of curves obtained by varying spatial location.

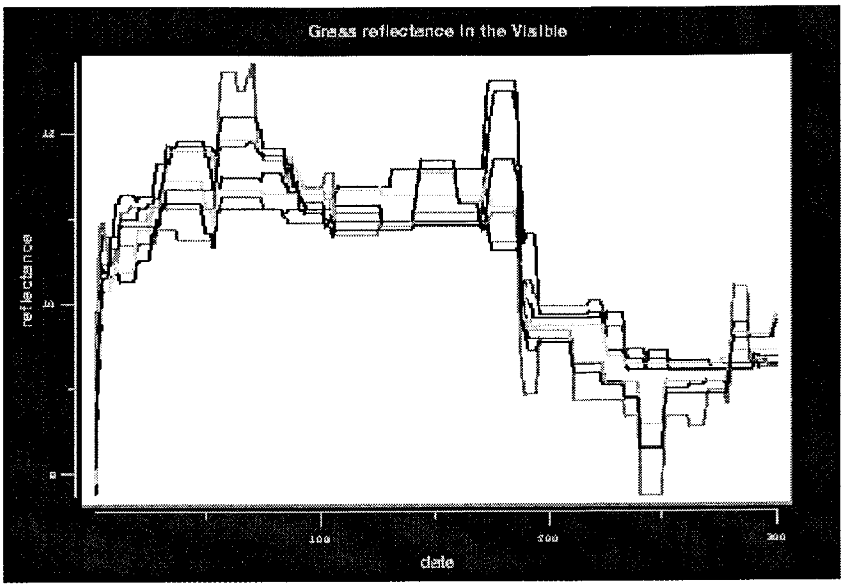


Figure 2 : Examples of curves of reflectance in the visible (grass).

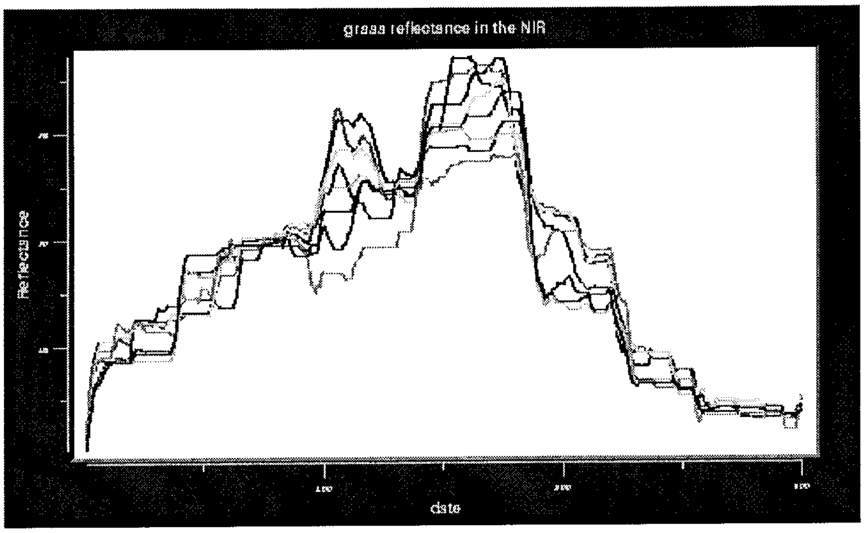


Figure 3 : Examples of curves of reflectance in the Near InfraRed (grass).

The second step is wavelet transformation and features extraction for curves coding. Among the computed curves we have chosen training samples (100 profiles) and validation samples (150 profiles). The training set is used to classes definition and the other set for evaluating the classification performance.

The classification method LDA has been applied into two spaces: the Principal Components space and the original features space. The table 1 presents the classification accuracy (on the test samples) for each land cover type:

Class	PCA and LDA	LDA
Grass	85%	87%
bare soil	61%	62%
Forest	52%	55%
Water	90%	92%

Table 1: Classification results.

We notice that we have better classification accuracy when applying LDA on the original feature vectors.

Both LDA and PCA project the original feature vectors onto a new space trough a transformation. The PCA changes both the location and the shape of the data in its transformed space whereas LDA only provides more class separability by building a decision between the classes. Obtaining a decrease of classification accuracy by PCA application before LDA can be explained by a loss of information after the first transformation.
The confusion matrix presented in table 2 gives more details about classification results.

	Grass	forest	Bare soil	water
grass	87%	0%	13%	0%
forest	8%	55%	21%	16%
bare soil	13%	24%	62%	1%
water	0%	8%	0%	92%

Table 2 :Confusion matrix.

It shows an important confusion between bare soil and forest. The reasons of this confusion can be investigated at two levels:

1- firstly, at the learning step: the training samples have to be carefully chosen, in such a way that they actually represent the land cover classes. In fact forest is a mixture of bare soil and trees, if we have a low density of trees the reflectance can be very close to that of bare soil.
2- secondly, at the feature extraction step: more relevant features to discriminate curves can be investigated in order to improve classification performance.

5. Scenario of application to land cover changes detection

The application of this work concerns land cover changes detection and especially erosion detection. The proposed framework is illustrated in figure 4.

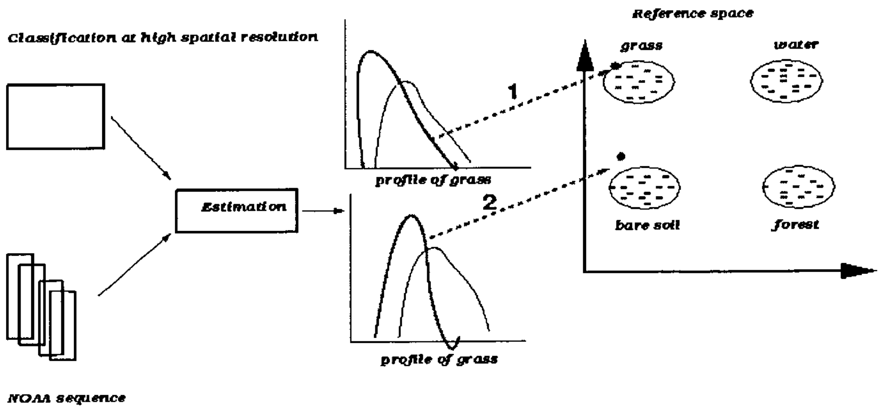


figure 4: Change detection scenario.

For a given site, we consider a year of reference for which we apply the approach described in section 2 to obtain the representation space of the land cover classes.

The next NOAA/AVHRR sequences provided after the reference year are used to estimate the new profiles for the land cover types of the same area. The obtained curves are then processed to extract relevant features then classified within the representation space corresponding to the reference year. Two different cases are possible for a profile representing locally a given land cover type:

1- The new profile is assigned to the corresponding land cover class. For example a profile assumed to be grass is classified as grass in the reference space. This implies that there is no change.
2- The new profile is closer to another land cover class. This implies that there is change and it has to be interpreted. For example a profile which is assumed to be grass is closer to the class representing bare soil. This may be interpreted by an erosion process occurring in the localization where the, profile is obtained.

6. Conclusion and perspectives

The work presented in this paper deals with a new application of signals classification. It experiments a process making it possible the detection of land cover changes at early stage.

The first obtained results are promising and we are investigating different ways to improve them. Improvements of classification accuracy can be expected by more careful experimentations to choose features for characterizing curves. We are studying parameters more relevant for the shape information such as the local extrema. We also plan to investigate the use of other wavelet families. Moreover, we are studying the method of wavelet optimization for classification proposed in [9]. It is an approach which allows the determination of the mother wavelet yielding the best classification result. It is realized by optimizing the filter coefficients according to a contrast criterion calculated on the learning set. An other way to improve classes discrimination proposed in [10, 11] can be also investigated on our data. It is a method to construct an orthonormal basis which maximize classes separability.

7. References

[1] H. Yang, R. Leich and R. Boite, «*Voiced speech coding at very low bit rates based on forward-backward waveform prediction* » IEEE trans on Speech Audio processing, vol 3 n°1.
[2] R. Kronland-Martinet, J. Morletand A. Grossman, « *Analysis of sound patterns through Wavelet Transform* » International Journal of Pattern Recognition and artificial Intelligence.
[3] K. Englehart, P.A. Hudings, B. Parker, and Stevenson M., « *Classification on the myoelectric signal using timefrequency based representation,* » Special issue of Medical Engineering and Physics on Intelligent data analysis in Bectromyography and Bectroneurography, 1999.
[4] S.G Mallat, « *Multifrequency channel decomposition of images and wavelt models,* » IEEE transaction on information theory, vol 37, no. 4, 1991.
[5] N. Intrator, Q.Q. Huynh, and Dobeck G.J., « *Feature extraction from acoustic backscattered signals using wavelet dictionnaries,* » in SPIE Wavelet Applications in Signal and Imaging processing, 1995.
[6] I. Daubechies, "*Ten lectures on wavelets,*" in CBMS-NSF Regional Conferences Series in applied Mathematics, 1992.
[7] NA. Quarmby, J.R.G. Townshend, J.J. Settle, K.H White, M. Milnes, T.L. Hindle, and N.Silleos, "*Linear mixture modelling applied to AVHRR data for crop area estimation*" International journal of remote sensing, vol. 13, 1992.
[8] S. Bouzidi, JP. Berroir, and Herlin I., "*A remote sensing data fusion approach to monitor agricultural areas*" in Proceedings of International Conference on Pattern Recognition, ICPR, 1998.
[9] M. Flucas, E. Doncarli, C. and Hitti, and Deschamps N., "*Wavelet optimization for classification*", in ICASSP, 1999.
[10] Saito N. and Coifman R.R., "*Local discriminant bases and their application*", J. Math, Imaging and Vision, vol 5, n° 4, pp. 337-358, 1995.
[11] R.R. Coifman and Wickerhauser M.V., "*Entropy based algorithms for best basis selections*" , IEEE Trans. Inform. Theory, vol 38, n° 2, pp. 713-719, 1992.