ADAPTIVE WAVELETS FOR IMAGE REPRESENTATION AND CLASSIFICATION

Gemma Piella, Marine Campedel and Béatrice Pesquet-Popescu

GET-ENST, Signal and Image Processing Department,
46 rue Barrault, 75634 Paris Cedex 13, France

ABSTRACT

In this paper we propose a feature-based wavelet representation for image classification and visualization. The work is primarily motivated by the need to classify quickly and efficiently large multispectral satellite images, and possibly to perform the classification task directly on compressed data. We propose a multiresolution approach based on a special class of adaptive wavelets which allows the extraction of salient features without loosing accuracy.

1. INTRODUCTION

Typically, classification involves feature extraction and representation, class modeling and parameter estimation. These steps are highly interdependent, since the choice of features influences the conditions under which a classifier operates and vice versa.

In fact, feature extraction is generally applicable to a wide range of imagery and tasks. It allows the identification of relevant features. Hence, the effective use of feature extraction can improve general analysis and interpretation of data. For classification, the goal of feature extraction is to produce a condensed information able to discriminate between different classes. The extracted attributes generally correspond to models produced by human experts; automatic selection procedures can also be added to reduce features redundancy or irrelevance.

Some of the most common feature extraction methods are principal component analysis, discriminant analysis feature extraction, and decision boundary feature extraction. These methods are usually referred to as statistic-based feature extraction [1]. In the context of image classification, typically extracted features correspond to textural and geometrical information. The textural part can be represented using classical wavelets based features (Gabor, QMF), statistics from co-occurrence matrices (Haralick coefficients [12]) or parameters estimated from Gaussian Markov Random Fields (GMRF [15]) models. The geometrical one traditionnaly comes from shape or contour analysis.

In the meantime, in the past two decades, the wavelet transform has been successfully applied in several applications, among which, feature extraction and classification. Due to their scale-space localization properties, wavelet transforms (and their variants such as wavelet packets) have proven to be an appropriate starting point for classification, specially for texture images, since classical wavelet transforms present some limitations when dealing with geometrical structures such as edges. Moreover, in [2, 3] it is sug-

gested that the choice of filters in the wavelet transform can be an important issue even for texture images.

In this paper, an adaptive wavelet transform [4] is applied to extract the physical features of remote sensing images. An algorithm based on non-linear approximation of this adaptive representation is used to identify significant features directly in the transformed domain. An important characteristic of our approach is that the transform preserves salient features so that the image can be analyzed and visualized at low resolutions while providing a compact representation. Moreover, the resulting decomposition can be directly used for nearly lossless compression of remore sensing images, thus allowing to jointly achieve efficient storing and browsing capabilities.

This paper is organized as follows. In the next section, we briefly describe the adaptive wavelet scheme used to represent the input images. In Section 3, we propose a feature extraction algorithm based on selecting the largest amplitude wavelet coefficients along the different scales and orientations as the important features for classification. In Section 4, we describe the classification method used in the simulations results, which are presented in Section 5. Concluding remarks are given in Section 6.

2. ADAPTIVE WAVELET REPRESENTATION

Our wavelet decomposition uses an adaptive update lifting step, followed by a fixed prediction step, such as illustrated in Fig. 1.
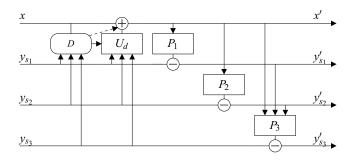

Figure 1: *2D adaptive wavelet decomposition.*

The input image bands x, y_{s_1}, y_{s_2} and y_{s_3} are the polyphase components of an original image x^0: $x(m,n) = x^0(2m, 2n)$, $y_{s_1}(m,n) = x^0(2m, 2n+1)$, $y_{s_2}(m,n) = x^0(2m+1, 2n)$, $y_{s_3}(m,n) = x^0(2m+1, 2n+1)$. The output x' corresponds to the *approximation* band, whereas $y'_{s_1}, y'_{s_2}, y'_{s_3}$ are the *detail* bands corresponding to different orientations (horizontal, vertical and diagonal, respectively).

In the update step, D is the so-called *decision map* which uses inputs from all four bands, and whose output is a binary

decision parameter $d \in \{0, 1\}$ which governs the choice of the update step. More precisely, if $d_\mathbf{n}$ is the output of D at location $\mathbf{n} = (m, n)$, then the updated value $x'(\mathbf{n})$ is given by

$$x'(\mathbf{n}) = \lambda_{d_\mathbf{n}} x(\mathbf{n}) + \sum_{j=1}^{J} \mu_{d_\mathbf{n},j} y_j(\mathbf{n}), \qquad (1)$$

with $y_j(\mathbf{n}) = y_{s_j}(\mathbf{n} + \mathbf{l}_j)$, $s_j \in \{s_1, s_2, s_3\}$, $\mathbf{l}_j \in L$. Here L is a window in $\mathbb{Z}^2$ centered around the origin. Note that the filter coefficients depend on the decision $d_\mathbf{n}$. In particular,

$$d_\mathbf{n} = D(\mathbf{v}(\mathbf{n})) = [p(\mathbf{v}(\mathbf{n})) > T],$$

where $[P]$ returns 1 if the predicate P is true, and 0 otherwise; p is a seminorm, T is a threshold and $\mathbf{v}(\mathbf{n}) \in \mathbb{R}^J$ is the gradient vector with components $v_j(\mathbf{n}) = x(\mathbf{n}) - y_j(\mathbf{n})$, $j = 1, \ldots, J$. Thus, our method builds lifting structures able to choose between two different update filters, the choice being triggered by the local gradient information of the input. If the gradient is large (in some seminorm sense) it chooses one filter, if it is small the other. We assume that the filter coefficients satisfy:

$$\lambda_d + \sum_{j=1}^{J} \mu_{d,j} = 1 \quad \text{and} \quad \lambda_d \neq 0, \quad \text{for } d = 0, 1.$$

It is easy to show that the gradient vector at synthesis $\mathbf{v}'(\mathbf{n}) \in \mathbb{R}^J$ with components $v'_j(\mathbf{n}) = x'(\mathbf{n}) - y_j(\mathbf{n})$, $j = 1, \ldots, J$, is related to $\mathbf{v}(\mathbf{n})$ by means of the linear relation $\mathbf{v}'(\mathbf{n}) = \mathbf{A}_d \mathbf{v}(\mathbf{n})$, where $\mathbf{A}_d = I - \mathbf{u}\mathbf{b}_d^T$, I is the $J \times J$ identity matrix, and $\mathbf{u} = (1, \ldots, 1)^T$, $\mathbf{b}_d = (\mu_{d,1}, \ldots, \mu_{d,J})^T$ are vectors of length J. The superindex 'T' denotes transposition. For brevity, we will henceforth suppress the argument $\mathbf{n}$ in our notation and write, e.g., x, y_j rather than $x(\mathbf{n}), y_j(\mathbf{n})$, respectively.

If $p(\mathbf{v}) \leq T$ at the analysis step, then the decision equals $d = 0$ and $\mathbf{v}' = A_0 \mathbf{v}$. If, on the other hand, $p(\mathbf{v}) > T$, then $d = 1$ and $\mathbf{v}' = A_1 \mathbf{v}$. We can have perfect reconstruction if we are able to recover the decision d from the gradient vector at synthesis $\mathbf{v}'$. For simplicity, we shall restrict ourselves to the case where d can be recovered by thresholding the seminorm $p(\mathbf{v}')$, i.e., the case that $d = [p(\mathbf{v}) > T] = [p(\mathbf{v}') > T']$, for some $T' > 0$. In [4] we have analyzed perfect reconstruction conditions for various seminorms p. An interesting case is the weighted ℓ^2-norm:

$$p(\mathbf{v}) = \left(\sum_{j=1}^{J} \omega_j v_j^2 \right)^{1/2}, \quad \text{with } \omega_j > 0. \qquad (2)$$

In this case,

C1: perfect reconstruction holds if $\mathbf{b}_0, \mathbf{b}_1$ are collinear with the weights $\omega = (\omega_1, \cdots, \omega_J)^T$, and $|\lambda_0| \leq 1 \leq |\lambda_1|$.

The combination of the adaptive update lifting with a fixed prediction lifting yields an adaptive wavelet decomposition step mapping x^0 into $x', \mathbf{y}' = \{y'_{s_1}, y'_{s_2}, y'_{s_3}\}$. We obtain an adaptive multiresolution wavelet decomposition by iterating such wavelet steps. That is, we can use the approximation x' as the input for another wavelet decomposition step and obtain $x'', \mathbf{y}''$. Then, we can repeat this for x'', and so on. Thus, iteration of K steps results in a K-level wavelet decomposition of x^0 into $\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^K, x^K$, where we have written $\mathbf{y}^1, \mathbf{y}^2$, etc., instead of $\mathbf{y}', \mathbf{y}''$, etc., for simplicity of notation.

3. FEATURE EXTRACTION

We wish to find regions in the image with high information content relevant to perform image classification. It is well-known that smooth image regions are represented by small wavelet coefficients, while edges and other singularities are represented by large coefficients. Moreover, wavelet transforms reflect the image structure in the following properties [5]:

(i) Large-magnitude coefficients tend to occur near each other within subbands (detail bands), and also at the same relative spatial locations in subbands at adjacent scales (level of resolution) and orientations.

(ii) Typical localized image structures (e.g., edges) tend to have substantial power across many scales at the same spatial location.

Thus, an appropriate way to extract the salient information is to detect and retain only the significant coefficients (whose magnitude is large enough). Moreover, since the assumption of subband independence was found valid for real images [6], we consider that $y'_{s_1}, y'_{s_2}, y'_{s_3}$ are independent.

We form the feature vector by applying the following simple algorithm:

Feature selection algorithm

1. Decompose the image with the adaptive wavelet scheme proposed in Section 2. We refer to this decomposition as AW (adaptive wavelets).
2. Keep the M-biggest coefficients (in magnitude) of the resulting decomposition. We refer to this selection as NLA (non-linear approximation).

In this paper, for the adaptive decomposition, we consider the horizontal and vertical neighbors of the sample to update, and use the weighted seminorm in (2) with $\omega_j = 1$ for $j = 1, \ldots, 4$. For $d = 0$, we take $\lambda_0 = 1/2$ and $\mathbf{b}_0 = \frac{1}{8}(1, 1, 1, 1)^T$, while for $d = 1$, $\lambda_1 = 1$ and $\mathbf{b}_1 = \mathbf{0}$ (i.e., no update is performed preserving thus the discontinuities). One can easily see that this choice satisfies the perfect reconstruction condition in C1. After the update step, a prediction step such as illustrated in Fig. 1 is performed. For simplicity, $P_1(x) = P_2(x) = x$ and $P_3(x, y_1, y_2) = y_1 + y_2 - x$.

Once the adaptive wavelet decomposition has been obtained, we keep the 3% highest magnitude coefficients. These coefficients form the feature vecture which will be used for the classification. For example, for an image of size 64×64, this means we retain 121 features.

Because of the use of the non-linear approximation (NLA) and adaptive wavelets (AW), we refer to our feature model as NLAAW.

4. CLASSIFICATION

Classification is performed to analyse the discriminative power of different feature sets. Four types of classifiers are used in our experiments: (1) kNN (k Nearest Neighbors) [16]; (2) Fisher classifier[1] [7]; (3) SVM[2] used with Gaussian kernel (with parameter 0.05) [8, 9]. Classification evaluation is performed using cross-validation. We typically use 5 validation loops because of our database size (100 data in each

---

[1] Implemented in Spider environment [10].
[2] SVM is derived from libsvm code (free sources[11]) using parameter C = 1000.

class, see section 5). The data set is randomly divided into 5 sets. One set is used to test the classifier obtained by training on the other 4 data sets. There are 5 loops of training and test, each data set being used only once for testing. The final result is the (%)mean±(%)standard deviation of the error rates obtained in each cross-validation loop.

5. EXPERIMENTAL RESULTS

The simulation results are based on small images (64x64) extracted from satellite panchromatic images (Spot 5, 5m/pixel), represented with 8 bits/pixel (256 gray levels). Originally the database contains 6 data classes with 100 examples for each class. The classes were manually constructed. Fig. 2 shows an example of each class.
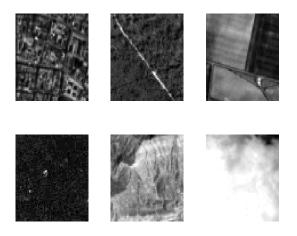

Figure 2: *6 texture classes, from left to right and top to down : city, forest, fields, sea, desert and clouds. (Copyright Centre National d'Etudes Spatiales)*

We compare our feature model (NLAAW) with three other feature models: Haralick [12] (13 different statistics computed on 4 oriented co-occurrence matrices, adding mean and standard deviation for each of them), Gabor [13] (mean and standard deviation in a 3-level decomposition with 4 orientations Gabor filters) and GMRF [15]. The GMRF parameters were recently introduced for satellite image indexing. They rely on a statistical estimation of Gibbs-Markov Random Fields model parameters: the first attribute corresponds to the parameters norm, the second to the fitting error variance while the third is the model evidence. The fourth one is related to the average gray level in each image. Note that we also consider the attributes VW that correspond to the variances computed in each subband in our wavelet decomposition.

In order to evaluate the effect of the adaptive filtering, we also introduce the features obtained when using non-adaptive wavelet filters (NLAW). In all wavelet-based cases two levels of decomposition were used.

Prior to classification, each feature is normalized to obtain mean 0 and variance 1. The results are displayed in Table 1. The last row shows the error rates obtained when combining the features of NLAAW with the variances of each subband in the adaptive decomposition.

In the second experiment, we consider only three classes (city, field and forest), which present a more geometrical

	D	kNN (k=8)	Fisher	SVM (rbf)
Haralick	78	13.8±2.2	30.8±4.3	7.0±1.5
Gabor	24	22.2±2.5	40.7±3.8	17.8±3.1
GMRF	4	9.7±4.5	16.8±3.3	12.5±2.6
VW	7	17.5±4.4	23.0±3.0	13.7±1.3
NLAW	121	7.8±2.3	**8.3±2.3**	6.0±0.9
NLAAW	121	7.0±2.5	9.7±2.5	5.8±0.8
NLAAW+VW	128	**5.3±1.8**	9.2±2.6	**3.0±1.8**

Table 1: *Classification error rates for different features sets using 6 texture classes. The dimension of each feature set is indicated by D column.*

structure (e.g., roads, lanes, boundaries, etc.) than the other three. The results are displayed in Table 2.

	D	kNN (k=8)	Fisher	SVM (rbf)
Haralick	78	9.3±3.3	22.0±5.2	4.7±2.2
Gabor	24	14.3±3.8	34.0±2.5	11.3±4.2
GMRF	4	5.0±2.6	16.7±4.7	4.7±1.8
NLAW	121	4.7±1.8	6.7±3.7	5.3±2.7
NLAAW	121	4.3±0.9	7.0±3.2	3.0±2.2
NLAAW+VW	128	**2.7±1.9**	**6.3±3.0**	**1.0±1.5**

Table 2: *Classification error rates for different features sets using 3 texture classes. The dimension of each feature set is indicated by D column.*

From these experiments we observe that:

- NLA based features provide the lowest error rates. The adaptive decomposition brings additional information when more geometrical structures are considered (see SVM better results when 3 classes are considered). Besides, we tested the algorithm with various AW decompositions, obtaining similar results. The choice of the parameters does not seem to influence very much. However, NLAW gives similar results when kNN and Fisher classifiers are considered, especially in the 6-classes problem.
- The lowest error rate is obtained with a combination of VW and NLAAW features. This indicates that VW provides a complementary information to NLAAW.
- The proposed features lead to much better classification results when used on the 3 classes with geometrical characteristics, where the adaptive decomposition can better take advantage of the input data.

It is important to mention that although the number of features is greater in the proposed approach, the complexity of our method is extremely low. When a subsequent wavelet compression method is applied to the sensed images, the extracted features come actually at no additional cost (neither computationally, nor in storage).

Moreover, when considering all features together (as a concatenated feature vector), we observe (see Table 3) that we do not enhance the classification performance. It is a well-known result from the feature selection literature that adding redundancy or noise cannot help the classification process. This means that the 128 NLAAW+VW features clearly contain all the discriminative information needed to classify our database.

	D	kppv(k=8)	fisher	svm(rbf 10)
All (6 classes)	234	5.0±1.3	8.7±1.5	**2.0±1.5**
All (3 classes)	234	2.7±1.9	6.0±3.0	**1.3±0.7**

Table 3: *Classification error rates for different features sets. The dimension of each feature set is indicated by D column. The first row corresponds to the classification on the whole set of classes, while the second row corresponds to the 3-classes problem.*

6. CONCLUSION AND FUTURE WORK

We have presented a simple and original feature extraction process based on adaptive wavelet decomposition called NLAAW (Non Linear Approximation and Adaptive Wavelets). This work is motivated by the need of an image representation allowing compression as well as classification. Our AW is introduced both to represent and analyze the image as well as to provide spatial-frequency based descriptors as features for classification.

The efficiency of the extracted attributes has been demonstrated through classification experiments performed on satellite images. In contrast with classical wavelets, our adaptive wavelet schemes are allowed to vary to fit data. Furthermore, the adaptive approach is able to reproduce the sharpness and smoothness of the irregular structures. This adaptive property guaranty the extraction of both geometrical and textural information contained in remotely sensed images, what cannot be performed by the classical feature extractors. Moreover the NLA scheme can be applied to traditional methods (like classical wavelets or block-coded based features) in order to enhance their discriminative power.

In conclusion, we have described some preliminary steps towards development of geometric and multiscale features to be used in classification of satellite images. Preliminary results show that we can obtain better class accuracy as compared to classical techniques, while performing only the necessary operations for compressing these images. As shown in another work [18], this adaptive representation also leads to better lossless and nearly lossless compression performance.

Concerning the feature set dimensionality, the rate of selected features (the largest 3% coefficients) in the NLA process was experimentally chosen and we are already convinced that the remaining features are still redundant. Hence we plan, as future work, to apply automatic feature selection process to the whole set of coefficients in order to determine the optimal number of features to be chosen. This kind of approach can also be used to compare the information brought by several feature sets [17].

All the presented experiments were performed on gray-level images. They can also be extended to color images and, more interesting for our application, to multispectral remote sensed images.

REFERENCES

[1] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, Academic Press, 1997.

[2] A. Mojsilovic, M. V. Popovic and D. M. Rackov, "On the selection of an optimal wavelet basis for texture characterization", *IEEE Transactions on Image Processing*, vol. 9, pp. 2043–2050, December 2000.

[3] M. Unser, "Texture classification and segmentation using wavelets frames", *IEEE Transactions on Image Processing*, vol. 4, pp. 1549–1560, November 1995.

[4] H. J. A. M. Heijmans, B. Pesquet-Popescu, and G. Piella, "Building nonredundant adaptive wavelets by update lifting", to appear in *Applied and Computational Harmonic Analysis*, 2005.

[5] E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain", in *Proceedings of SPIE 44th Annual Meeting*, vol. 3813, pp. 188–195, Denver, Colorado, July 1999.

[6] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models", *IEEE Transactions on Image Processing*, vol. 10, pp. 1056–1068, July 2001.

[7] T. Hastie, R. Tibshirani and J. Friedman, "The elements of statistical learning", *Springer Series in Statistics, Springer-Verlag*, 2001.

[8] B. Schoelkopf and A. Smola, "Learning with Kernels", *MIT Press*, 2002.

[9] C. Cortes and V. Vapnik, "Support-vector networks", *Machine Learning*, 20(3), pp. 273-297, 1995.

[10] J. Weston, A. Elisseeff, G. Bakir and F. Sinz, "The Spider for Matlab - v1.4", *Max Planck Institute for Biological Cybernetics*, 2004.

[11] C.C. Chang and C.J. Lin, "Libsvm 2.5", *http://www.csie.ntu.edu.tw/~cjlin/libsvm*, 2004.

[12] R. M. Haralick, "Statistical and structural approaches to texture", *Proceedings of IEEE*, vol. 67, pp. 786-904, May 1979.

[13] I. Fogel and D. Sagi, "Gabor filters as texture discriminator", *Biological cybernetics*, vol. 61, PP. 103-113, 1989.

[14] R. Chellappa, "Two-dimensional discrete Gaussian Markov random field models for image processing", *Pattern Recognition*, vol. 2, pp. 79-112, 1985.

[15] M. Datcu, K. Seidel and M. Walessa, "Spatial information retrieval from remote-sensing images. I. Information theoretical perspective",*IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no 5, pp 1431-1445, 1998.

[16] P.E. Hart, "The condensed Nearest Neighbor Rule", *IRE Transactions on Information Theory*, col. IT, no 14, pp 515–516, 1968.

[17] M. Campedel, B. Luo, H. Maitre, E. Moulines, M. Roux and I. Kyrgyzov, "Indexation des images satellitaires : Détection et évaluation des caractéristiques de classification", *technical report ENST 2004D008*, 2004.

[18] G. Piella, B. Pesquet-Popescu, H. Heijmans, "Adaptive Wavelets by Update Lifting", to appear in the Special Issue of "Signal Processing: Image Communication" on ' 'European Projects on Visual Representation Systems and Services", 2005