A New Dimensionality Reduction Method for Seabed Characterization: Supervised Curvilinear Component Analysis

Hicham Lanaaya*§, Arnaud Martin§, Driss Aboutajdine*, Ali Khenchaf§
*Faculté des sciences de Rabat
Avenue Ibn Batouta, B.P. 1014
Rabat, Morocco
Email: aboutaj@fsr.ac.ma
§ENSIETA / E3I2 EA3876
2, rue François Verny
29806, Brest Cedex 9, France
Email: {laanayhi, Arnaud.Martin, Ali.Khenchaf}@ensieta.fr

***Abstract*— In this paper, we present a new method for dimensionality reduction, called supervised Curvilinear Component Analysis, for the classification of sonar images task using support vector machines. Indeed it is important in many underwater applications to get tools that give automatically the kind of sediments. This method derives from the known method Curvilinear Component Analysis. It gives good results for data not highly overlapped. We have used this method after a feature extraction step based on wavelet decomposition applied to our sonar images database.**

I. Introduction

The sonar imaging is one of the advanced methods for data acquisition about of sea floor. Detecting a kind of sediment can be important. For example the rocks can be used as land-marks for images registration being used for underwater navigation, or for the creation of underwater map used by the sedimentologists. A skilled expert can interpret the images of the surveyed area and produce a base map showing the distribution of different classes of sediments.

To perform the sonar images classification, we adopt the Knowledge Discovery on Database (KDD) process as shown in Fig. 1. There are four principal steps for the KDD process, given a database, the preprocessing step gives us sonar images which represented by the grey level of the pixels in the image that correspond to the acoustic reflectance. Then we do a feature extraction to extract the relevant features. In most cases, we have to reduce the number of extracted features, that is the dimensionality reduction step which is the purpose of this paper. The dimensionality reduction allows us to perform the classification task in a low dimensional space: that is the last step for the KDD process.

There are different methods cited in [1] of feature extraction for image processing. These methods are quite similar, here the wavelet decomposition is choosen.

Classification step has been also studied in our previous works [2], [3], [4]. We propose here the use of a supervised

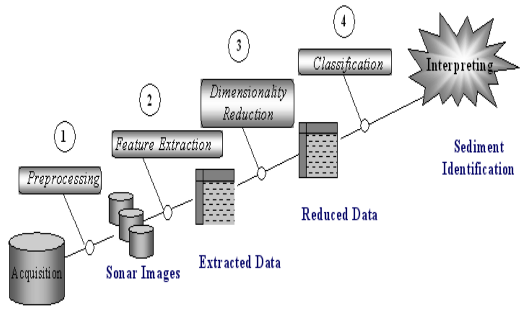


Fig. 1. The process KDD for sonar images classification.

classification: the Support Vector Machines (SVM). This approach has proved its performance in the case of nonlinear data.

In a previous paper [3] we have studied the problem of dimensionality reduction by features selection using genetic algorithms, that means keep the best previous extracted features considering that genetic algorithms use a feedback between the classification task and the selection. This method aims finding these best features while maximizing the classification rate. Here, we want to reduce the dimension while keeping the maximum of information. Most common methods try to conserve inertias of data by the variations. Most of these methods are linear such as principal component analysis (PCA) which is unsupervised and linear discriminant analysis (LDA) which is supervised. Here, a nonlinear dimensionality reduction method called Curvilinear Component Analysis (CCA, [5]) is considered in order to reduce the dimension of data, as shown in [2]. CCA reduces the time elapsed during the classification but it can not give a better classification rate because it does not take into account the class of points.

It is why we have tried to extend the concept of CCA to multiple manifolds or classes, each representing data of one specific class, a supervised variants of CCA is proposed for dimensionality reduction and to increase the classification rate. In this paper, a framework unifying the unsupervised and supervised methods is given. Supervised CCA is then applied to an artificial database and then on our sonar images database and is shown to be useful for high-dimensional data with a clear manifold structure where the classes are not highly overlapped.

We present the principle of CCA and the proposed supervised CCA in section II. In section III we present the classification step based on SVM. Finally, in section IV, experimental results of classification following the KDD process are proposed on generated data and on a real sonar images database where the feature extraction based on wavelet is recalled.

II. Curvilinear Component Analysis framework

A. CCA

The distance between various points of a set of individuals plays a significant role in classification. Thus, preserving the same topology of the input data in a low-dimension space will enable to gain in time of classification.

To deal with the problem of high-dimensional space in classification task, we use dimensionality reduction methods. Here, we introduce CCA a nonlinear dimensionality reduction method. It consists to preserve the local topology on the contrary of the PCA which is a linear method which seeks to maximize the standard deviation. CCA has been already presented in a number of works [5], [2]: CCA takes a set of N D-dimensional vectors (x_i) as input and maps them to a set of N M-dimensional, (y_i) vectors where the y_i corresponds to x_i and $M << D$, while preserving the local topology. In the CCA, the topology is defined by the distances between all pairs of vectors of original data. Since the topology cannot be entirely reproduced in the projection subspace, which has a lower dimension than the original subspace, the local topology, the most important, is favored to the detriment of the global topology. The goal of CCA is then to minimize an error function which characterizes the difference of topology between the original subspace (x_i) and the projection subspace (y_i):

$$E = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(X_{ij} - Y_{ij})^2 F_\lambda(Y_{ij}). \qquad (1)$$

with: X_{ij}: represent the euclidean distance between the inputs x_i and x_j in $\mathbb{R}^D$.
Y_{ij}: euclidean distance, between the projections y_i and y_j of the inputs x_i and x_j in the projection subspace $\mathbb{R}^M$.
$F : \mathbb{R}^+ \rightarrow [0, 1]$ is a decreasing function of its argument, so it is used to favor local topology preservation. For example, F could be a step, exponential or sigmoïd function of Y_{ij}.

To get the outputs y_i a random point y_i is choosen, and all the $y_{j\neq i}$ are moved with respect to y_i with the rule:

$$\forall j \neq i \qquad \Delta y_j = \alpha(t) f_\lambda(Y_{ij})(X_{ij} - Y_{ij})\frac{y_j - y_i}{Y_{ij}}, \qquad (2)$$

where $\alpha(t)$ is an adaptive factor that evolves with time, the complexity is in $O(N)$. CCA allows a reduction of size without decreasing classification performances. In the following section we introduce supervised CCA in order to reduce the dimensionality and increase classification rates.

B. Supervised CCA

We introduced the notion of Supervised CCA to treat data knowing the class of each individual and so outperforms the classification rates. We search to preserve the local topology knowing the class of each point, on the contrary of LDA, that is a linear method where we maximize the covariance between-class and minimize the covariance within-class.

Consequently we have modified the CCA algorithm to take into account the class of individuals, thus the algorithm of the supervised CCA will be as follow:
suppose that we have to represent variables $x_i \in \mathbb{R}^D$ (the inputs) with variables $y_i \in \mathbb{R}^M$ (the outputs).
The new algorithm is given by:

Initialization of y_i
Initialization $t = 0$
For each t
˜ *Evaluate* $\alpha(t)$ *and* $\lambda(t)$
˜ *For each individuals* y_i
˜ $\Delta y_j = \alpha(t) F_{\lambda(t)}(Y_{ij})(X_{ij} - kY_{ij})\frac{y_j - y_i}{kY_{ij}}$
˜ *with* $j \neq i$ *and Class(*x_i*)=Class(*x_j*)*
˜ *End For*
End For

The k used controls the level of regrouping of the individuals of each class. In this approach, for each y_i, one moves y_j such as $i \neq j$ and y_i, y_j belong to the same class. The problem is how to represent new vectors which are not with the x_i used. For that, being given a new vector x_0, to find the y_0, we will minimize the following error by using the gradient descent method:

$$E_0 = \frac{1}{2}\sum_{i}\sum_{j\neq i}(X_{0j} - X_{0j})^2 f_\lambda(Y_{0j}). \qquad (3)$$

Thus, instead of moving each vector with respect to each other, only one point is adapted according to a simple stochastic gradient descent, while all the others are kept fixed. Therefore, this point is searched with respect to the outputs y_i with respect to the measured distances X_{i0}. It is actually a local mapping and the initialization of y_0 is made randomly.

III. Support Vector Machines

In the classification task, the images are analyzed in order to be separated. This process uses some features of the images to differentiate every one from the others. This way, the images can be classified in several classes with some characteristic in common. Then the classification of sediments can be done using anyone of well-known classification techniques. One of them is a supervised method called SVM given a simple way to obtain good classification results with a reduced knowledge. So, the used classification is based on the SVMs classification algorithm. The principle of SVMs has been developed by Vapnik [6] and used in several applications [4], [7]. The classification task is reduced to find a decision border dividing the data into groups representing the separated classes. The simplest decision case is when the data can be divided into two groups. Consider the problem where the vectors can be divided into two sets. We must find the optimal decision border that separates these two sets of images. This optimal election will be the one that maximizes the distance from the border to the data. In the two dimensional case, the border will be a line, and in a multidimensional space the border will be an hyperplane (*cf* Fig. 2). The searched decision function is given by:

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i < x_i, x > +b. \qquad (4)$$

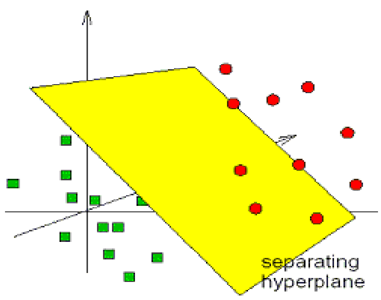


Fig. 2. An separate hyperplane for two classes in 3 dimension.

The y values of this expression are +1 for positive classification training vectors (representing one class) and -1 for the negative training vectors (representing the other class). Also, the inner product is performed between each training input and the vector which must be classified. Thus, we need a set of training data (x, y) in order to find the classification function. The values α_i are the Lagrange multipliers, b a constant value obtained by the minimization process and l the number of vectors in the training database. These vectors with a value different to zero, are known as support vectors. In our case, x represents one image from the sonar images training database (in the space of features) and y represents the predicted kind of sediment present on the x image. The (x_i, y_i) represents the images of the training database and there associated kind of sediments. When data are not linearly separable this scheme cannot be used directly. To avoid this problem, the SVMs map the input data into a high dimensional features space. The SVM constructs an optimal hyperplane in the high dimensional space and then returns to the original space transforming this hyperplane in a nonlinear decision border. The nonlinear expression for the classification function is given in the following equation:

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b, \qquad (5)$$

where K is the kernel that performs the nonlinear mapping. The choice of this nonlinear mapping function or kernel is very important in order to obtain good classification performance. But, there are no method to do this choice. The first kernel investigated were the following:

- Linear $K(x, y) =< x, y >$,
- Polynomial $K(x, y) = (< x, y > +1)^p)$,
- Gaussian $K(x, y) = exp(-\gamma(x - y)^2)$, where γ is a parameter that will be tuned by the user.

When some data into the sets cannot be separated, the SVM can include a penalty term, C, in the minimization problem, which makes more or less important the misclassification. The greater is this parameter, the more important is the misclassification error into the minimization procedure if classes are not overlapped.

This approach can be generalized to more then two classes [8], [9] where we can quote different methods:

- The direct approach, where we considered directly all the classes,
- One-vs-rest: we made a classifier for each two classes, and then we fuse the results,
- One-vs-one: we seek to separate each class from the others, and then we fuse the results,

The direct approach is a straightforward generalization of the support vector concept to more than two classes. The pairwise one-vs-one method conserves most of the maximal margin hyperplanes, but the simple one-vs-rest scheme coincides only in single points with the constructed optimal separating hyperplanes. It is why we prefer the one-vs-one approach.

IV. Experiments

The SVM classifier used on our experiments is *libsvm* given in [10]. This algorithm use the one-vs-one multi-class approach. The SVM classifier was trained using the training database. We made our experiments on two databases, an artificial one and a real one from sonar images.

A. Artificial data

The first example is a 1000-points synthetic dataset of 6 gaussians in 15-dim with covariances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2 = 0.25$ and means are choosed in order to obtain quite separated classes, see Fig. 3. To visualize the data points, we have projected them to the two first components, using PCA, which gives the representation with maximum variation.

We have divided the artificial data into two equitable databases, one for learning task and the other one for testing. Thus, each class of the learning database contains 500 15-dimensional points. For tests, we have used a linear kernel, this kernel gives best results in such data, for SVM classification with $C = 1$ the default value of $libsvm$ classifier. We have used 5 as dimension of the output dimension of both CCA and supervised CCA ($D = 15$ and $M = 5$) in order to get the same dimension as in our sonar images database tests. The results of SVM classification are shown in Table I, II and III.

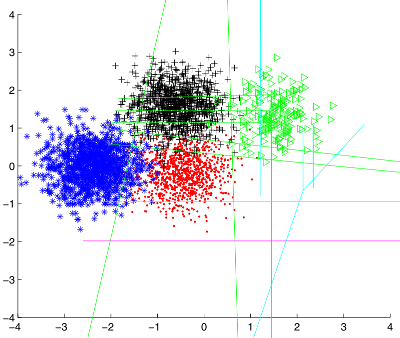


Fig. 4. Example of sonar image (provided by GESMA group).

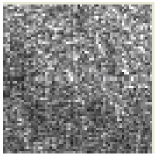
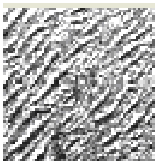
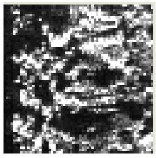
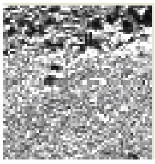
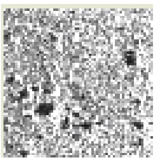
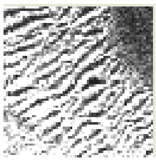
 1. Sand	 2. Ripple	 3. Rock
 4. Rock & Sand	 5. Cobbles	 6. Ripple & Sand

TABLE IV
EXAMPLE OF SMALL IMAGES

parameters. The energy is given by:

$$\frac{1}{HL}\sum_{n=1}^{H}\sum_{m=1}^{L} I_d^i(n,m), \quad (6)$$

where H and L are respectively the number of pixels on the rows, and on the columns. The entropy is estimated by:

$$-\frac{1}{HL}\sum_{n=1}^{H}\sum_{m=1}^{L} I_d^i(n,m)\ln(I_d^i(n,m)), \quad (7)$$

and the mean is given by:

$$\frac{1}{HL}\sum_{n=1}^{H}\sum_{m=1}^{L} |I_d^i(n,m)| \quad (8)$$

So we obtain 15 features. Each small-image is then represented in a 15-dimension space.

	Sand	Rock	Shadow	Ripple	Silt	Cobbles	Total
Tr. DB	971	319	79	147	23	18	1557
Ts. DB	1350	596	293	227	211	15	2692

TABLE V
TRAINING AND TEST DATABASE EFFECTIVE.

3) Results: On the Table VI, we present the obtained effective of each kind of sediment after the tests made on our test database without dimensionality reduction. Experiments are made on the sonar image database after a wavelet decomposition, the features dimension is 15 ($D = 15$) an the output dimension is 5. In all experiments, we have used a SVM classifier with gaussian kernel with $\gamma = 0.0404$ and $C = 100$, parameters that gives the best classification rates.

We have obtained a global classification rate of 67.57% defined as the number of good classified small-images on the total of small images. Notice that no cobbles small-images are detected. 1122 of 1350 (83.11%) of the sand small-images are detected, 73.32% of the rock small-images are well classified and 51.53% of the shadow small-images are detected. We note a low rate of detection for the two sediments, silt and ripple; indeed, only 43.80% (respectively 13.27%) of the ripple (respectively silt) small-images are detected. The classifier tends to classify all the images in the two classes, sand and rock small-images, both majority classes of the database in terms of effective. Before training the classifier, we apply CCA

		References					
	Class name	Sand	Rock	Shadow	Ripple	Silt	Cobbles
Test	Sand	1122	108	101	9	10	0
	Rock	51	437	80	20	8	0
	Shadow	124	17	151	1	0	0
	Ripple	98	42	8	79	0	0
	Silt	53	42	88	0	28	0
	Cobbles	10	5	0	0	0	0

TABLE VI
CONFUSION MATRIX FOR SONAR IMAGES DATA WITHOUT DIMENSIONALITY REDUCTION (CLASSIFICATION RATE 67.57%)

for dimensionality reduction, classification results are shown on the Table VII We have obtained a global classification rate of 53.42%. 84.75% of sand small-images are well classified a rate rather than on the rough data, and 44.29% of rock small-images are detected. No cobbles, silt and ripple small-images are detected. We obtained a classification rate of 11.26% for shadow small-images. Thus, we have obtained a classification rate less than the classification rate obtained on our sonar database without dimensionality reduction. On the Table VIII we give results after applying supervised CCA on the sonar image database. We obtained a classification rate of 48.55%, a rate lower than the classification rate obtained by applying SVM on our rough database. One gained in computing time but one lost on the classification rate; theses results can be explained by the fact that the six classes of our sonar database are highly overlapped and the classes are unbalanced.

	Class name	References Sand	Rock	Shadow	Ripple	Silt	Cobbles
Test	Sand	1141	149	60	0	0	0
	Rock	264	264	68	0	0	0
	Shadow	243	17	33	0	0	0
	Ripple	176	51	0	0	0	0
	Silt	25	91	95	0	0	0
	Cobbles	11	4	0	0	0	0

TABLE VII

CONFUSION MATRIX FOR SONAR IMAGES DATA AFTER CCA (CLASSIFICATION RATE 53.42%)

	Class name	References Sand	Rock	Shadow	Ripple	Silt	Cobbles
Test	Sand	1080	121	0	145	2	2
	Rock	281	192	0	112	0	11
	Shadow	244	45	0	4	0	0
	Ripple	173	21	0	33	0	0
	Silt	20	149	0	28	2	12
	Cobbles	11	1	0	3	0	0

TABLE VIII

CONFUSION MATRIX FOR SONAR IMAGES DATA AFTER APPLYING SUPERVISED CCA (CLASSIFICATION RATE 48.55%)

Note that with supervised CCA we detect some ripple, silt and cobbles small-images, but no shadow small-images, on the contrary of the CCA.

V. CONCLUSION

The dimensionality reduction is a necessary preprocessing step for classification of sonar images data following the KDD process. In this paper, we have presented and used a new method for dimensionality reduction in the context of sediment classification. We have shown that the application of supervised CCA gives better results in the artificial data used for experiments. However results are not so concluding on our real sonar images database. We can explain this difference by the fact that the classes are highly overlapped and unbalanced.

In our experiments, we have used an heuristic value of the output space of supervised CCA, further works will focus on adding an automatic tuning of this parameter by searching the intrinsic dimension, the small dimension where we can represent data without lose of information.

The data used for our test are unbalanced, a thing that have a negative effect in classification task as shown in experiments part.

Another problem is the patch-worked small-images. We are working on the realization of a new repartition of the data with a previous manual segmentation of the sediment.

REFERENCES

[1] A. Martin, G. Svellec, and I. Leblond, "Characteristics vs decision fusion for sea-bottom characterization," *Colloque Caractérisation in-situ des fonds marins, Brest, France*, 21-22, October 2004.

[2] H. Laanaya, A. Martin, A. Khenchaf, and D. Aboutajdine, "Dimensionality reduction of sonar images for sediments classification," *Colloque Caractérisation in-situ des fonds marins, Brest, France*, 21-22, October 2004.

[3] ——, "Feature selection using genetic algorithm for sonar images classification with support vector machines," *European Conference on Propagation and Systems, Brest, France*, vol. 24(11), 15-18, March 2005.

[4] C. Archaux, H. Laanaya, A. Martin, and A. Khenchaf, "An svm based churn detector in prepaid mobile telephony," *International Conference On Information & Communication Technologies (ICTTA), Damas, Syrie*, pp. 19–23, 2004.

[5] P. Demartines and J. Hérault, "Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets," *IEEE Transaction on Neural Networks*, vol. 8(1), pp. 711–720, 1998.

[6] V. N. Vapnik, *Statistical Learning Theory*. John Wiley and Sons, 1998.

[7] H. Laanaya, A. Martin, A. Khenchaf, and D. Aboutajdine, "Feature selection using genetic algorithm for sonar images classification with support vector machines," *European Conference on Propagation and Systems, Brest, France, 15-18 March*, 2005.

[8] J. Weston and C. Watkins, "Multi-class support vector machines," 1998. [Online]. Available: citeseer.ist.psu.edu/8884.html

[9] C. W. Hsu and C. J. Lin, "A comparison of methods for multi-class support vector machines," *Technical report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan*, 2001.

[10] C. C. Chang and C. J. Lin, "Libsvm: a library for support vector machines," *Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm*, 2001.