

# Characteristics versus decisions fusion for sea-bottom characterization

Arnaud MARTIN, Gwénola SEVELLEC,  
Isabelle LEBLOND  
ENSIETA/E<sup>3</sup>I<sup>2</sup>, EA 3876  
2, rue F. Verny,  
29806 Brest, Cedex, France  
Tel.: +(33)2.98.34.88.84  
Arnaud.Martin@ensieta.fr

**Abstract** – *The automatic sea-bottom characterization is a difficult problem. The most of automatic characterization approaches are based on texture analysis. Indeed, the sonar sea-bottom images present many homogeneous areas of sediment that can be seen like a sonar texture. However, texture characterization approaches allow different results according to the kind of texture. In order to improve the sediment classification, we propose to use two information fusion approaches coming from the evidence theory. One estimate the belief function with a probabilistic point of view (decision level), the other one with a distance approach (characteristic level).*

**Keywords:** sea-bottom characterization, decision fusion, characteristic fusion.

## 1 Introduction

The sea-bottom characterization is a difficult problem. The automatic characterization approaches are based on texture analysis; the sonar sea-bottom images present homogeneous or not part of sediment that can be seen like a sonar texture [1]. The state of art presents many techniques for texture analysis and the choice of one or more of them depends, the most of time, of the kind of images and the application. From these techniques a reduce number of relevant feature is calculated in order to classify the images in one or more sediment types.

In this paper, in order to outperform the sea-bottom characterization, we propose to fuse the data at the level of the characteristics (*i.e.* the numerical outputs of the classifiers) or at the level of the classification results (*i.e.* at the level of the decisions of the classifiers).

To do this, we follow the process schema presented in Fig. 1. We present the database of sonar images in section 5. We consider four methods for texture characterization: the co-occurrence matrices, the run-lengths matrices, a wavelet transform describe in [2] and Gabor filters, presented in section 2. A neuronal network is used to classify extracted features into each methods of texture characterization (see section 3). Results obtained with each approach are different: the run-

lengths method provides the worst classification rate whereas the co-occurrence matrices provide the best global classification rate. But the results are also different according to the type of sediments. Indeed, for example, the co-occurrence matrices technique is not invariant in rotation: the ripple sediment is misclassified. The wavelet transform and Gabor filters studied techniques are invariant in rotation, but provide on isotropic texture classification rates weaker than the co-occurrence matrices.

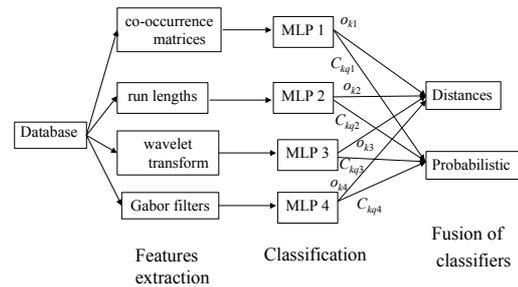


Fig. 1. Process schema.

The interest of information fusion is to take into account the imperfections of the considered techniques in order to perform classification rates. We propose to study some approaches coming from the evidence theory, estimating the belief function based on a probabilistic point of view, and based on distance, applied respectively for the decision fusion (*i.e.* the decisions of each classifier, noted  $C_{kq}$  on Fig; 1) and for the characteristic fusion (*i.e.* the outputs of each classifier, noted  $o_k$  on Fig; 1). We present these both approaches in section 4.

Experimental results are given in section 6.

## 2 Feature Extraction

Following the process presented on Fig. 1, four common texture characterization approaches are used. We recall here their principle in [2] and [3].

## 2.1 Co-occurrence matrices

The co-occurrence matrices are calculated, by numbering the occurrence of identical grey level of two pixels. We consider here four directions:  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  and  $135^\circ$ . In these four directions, six parameters given by Haralick [4] are calculated.

The first parameter characterize the homogeneity:

$$\sum_{i=1}^{n_g} \sum_{j=1}^{n_g} c_d^2(i, j), \quad (1)$$

where  $n_g$  is the number of grey levels and  $c_d(i, j)$  is the estimation of the probability of transition of the pixel  $i$  to the pixel  $j$  in the direction  $d$ . Contrast estimation is given by:

$$\frac{1}{n_g - 1} \sum_{k=0}^{n_g-1} k^2 \sum_{i, j: |i-j|=k} c_d(i, j). \quad (2)$$

The entropy is estimated by:

$$1 - \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} c_d(i, j) \ln(c_d(i, j)). \quad (3)$$

The correlation is given by:

$$\frac{1}{\sigma_x \sigma_y} \left| \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} (i - m_x)(j - m_y) c_d(i, j) \right|, \quad (4)$$

where  $m_x$  and  $m_y$  describe the mean on rows and columns of  $c_d$  respectively, and  $\sigma_x$  and  $\sigma_y$  are the standard deviations. The directivity is given by:

$$\sum_{i=1}^{n_g} c_d(i, j), \quad (5)$$

and the uniformity by:

$$\sum_{i=1}^{n_g} c_d^2(i, j). \quad (6)$$

This classical approach allows parameters. The problem of the co-occurrence matrices is the non-invariance in translation. Typically, the problem can appear for the ripple texture characterization.

## 2.2 Run-lengths method

The run-lengths matrix is obtained by counting consecutive pixels with the same grey level in the four previous directions. Hence a matrix  $\mathbf{L}_d = (L_d(i, j))$  is obtained, where  $L_d(i, j)$  is the number of run lengths  $j$  of the pixels with a grey

level  $i$  in the direction  $d$ . So the number of run lengths is given by:

$$N_L = \sum_{i=0}^{n_g-1} \sum_{j=1}^{n_g} L_d(i, j). \quad (7)$$

Then five parameters are extracted from the four directional matrices. The first one is the proportion of small run-lengths given by:

$$\frac{1}{N_L} \sum_{i=0}^{n_g-1} \sum_{j=1}^{n_g} \frac{L_d(i, j)}{j^2}. \quad (8)$$

The second one is the proportion of big run-lengths given by:

$$\frac{1}{N_L} \sum_{i=0}^{n_g-1} \sum_{j=1}^{n_g} j^2 L_d(i, j). \quad (9)$$

The run dispersion between the grey levels is given by:

$$\frac{1}{N_L} \sum_{i=0}^{n_g-1} \left( \sum_{j=1}^{n_g} L_d(i, j) \right)^2, \quad (10)$$

and the run dispersion between the lengths is:

$$\frac{1}{N_L} \sum_{j=1}^{n_g} \left( \sum_{i=0}^{n_g-1} L_d(i, j) \right)^2. \quad (11)$$

We can also use the length percentage  $\frac{N_L}{N}$  where  $N$  represents the number of pixels.

This method is well suited in the case of optical image for example, where no speckle is present. Anyway, in the case of sonar images, we have to remove first the speckle or adapt the parameter calculation. However, we keep this approach in order to study the effect of a bad extraction of texture features.

## 2.3 Wavelet transform

The both previous approaches do not consider the translation invariance in the directions. The discrete translation invariant wavelet transform is based on the choice of the optimal translation for each decomposition level [2]. Each decomposition level  $d$  gives four new images. We choose here a decomposition level  $d=3$ . For each image  $I_d^i$  (the  $i^{\text{th}}$  image of the decomposition  $d$ ) we calculate three parameters. The energy is given by:

$$\frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M I_d^i(n, m), \quad (12)$$

where  $N$  and  $M$  are respectively the number of pixels on the rows, and on the columns. The entropy is estimated by:

$$-\frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M I_d^i(n, m) \ln(I_d^i(n, m)), \quad (13)$$

and the mean is given by:

$$\frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M |I_d^i(n, m)|. \quad (14)$$

So we obtain 63 wavelet features ( $3+4*3+16*3$ ).

## 2.4 Gabor filters

The impulsional response of the Gabor filter is given by:

$$\exp\left[-\frac{1}{2}\left(\frac{f^2(x-x_0)}{\sigma_x^2} + \frac{g^2(y-y_0)}{\sigma_y^2}\right)\right] * \cos(2\pi u_0 x + \varphi)$$

where  $u_0$  is the radial frequency of the filter,  $\sigma_x$  and  $\sigma_y$  are the standard deviations,  $\varphi$  is the phase,  $(x_0, y_0)$  corresponds to the point where the Gaussian is maximum, and where:

$$\begin{cases} f(x-x_0) = (x-x_0) \cos \theta + (y-y_0) \sin \theta \\ g(y-y_0) = -(x-x_0) \sin \theta + (y-y_0) \cos \theta \end{cases} \quad (15)$$

and  $\theta$  is the rotation angle. We consider five different frequencies and six directions, that is 30 filters are designed. Then we calculate four parameters as in [3]. The first one is the maximum value of the matrix numbers normalized by the mean, which represents the maximum value of the standard deviation with the considered sediment. The mean of all points of the matrix is also calculated. The third parameter represents the mean on the horizontal directions only (pings directions) normalized by the global mean. The last one is the global standard deviation before filtering.

This approach takes the translation invariant on the directions into account.

## 3 Multilayer perceptron classifier

The multilayer perceptron (MLP) is a feed-forward fully connected neural network [5, 6].

The data  $x$  is described by  $n$  parameters  $(x_1, \dots, x_n)$ . Each unit of the network is an artificial neuron (perceptron) with the structure shown on Fig 2.

All the unit outputs of each layer are connected to all the unit inputs of the next layer weighted by the values  $w_{lj}$ , where  $l$  is the source unit and  $j$  is the target unit. These weights are initialized with small

random values and they reach stable values after the learning process if it converges.

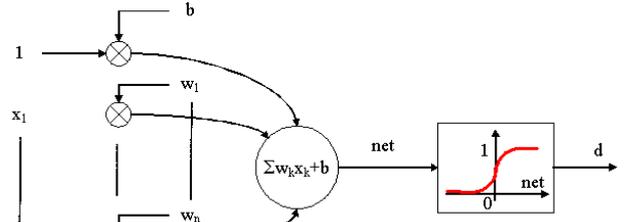


Fig. 2. Artificial neuron structure.

The learning process consists in repeated presentations of the training vector and of the corresponding desired output vector to the network. The objective of the learning process is to minimize the quadratic error:

$$\varepsilon = \frac{1}{2} \sum_{j=1}^m (d_j - o_j)^2 \quad (16)$$

where  $o_j$  represent the real output of the unit  $j$  of the last layer.

If the sigmoid function shown on Fig. 2 is used then the following learning algorithm is obtained:

$$\begin{cases} w_{lj}(t+1) = w_{lj}(t) + \eta \delta_j(t) o_l(t) \\ \delta_j = c o_j [1 - o_j] [d_j - o_j], \text{ for the output layer} \\ \delta_j = c o_j [1 - o_j] \sum_l \delta_l w_{lj}, \text{ elsewhere} \end{cases}$$

The constant  $c$  controls the slope of the sigmoid function and  $\eta$  stands for the learning rate.

This rule is known as the back propagation algorithm or the generalized delta rule. Its convergence can be improved if a momentum term is added and if the learning rate is tuned in an appropriate manner [6].

## 4 Fusion models

The evidence theory allows for a representation of both imprecision and uncertainty through two functions: plausibility and belief [7, 8]. Both functions are derived from a mass function defined on each subset of the space of discernment  $D = \{C_1, \dots, C_m\}$  onto  $[0, 1]$ , such that:

$$\sum_{A \subseteq D} m(A) = 1, \quad (17)$$

where  $m(\cdot)$  represents the mass function.

The first difficulty is the choice of a mass function. There are two types of approaches: one based on a probabilistic model [8] and another one based on distance transformation [9]. Appriou [8]

proposes two equivalent models based on three axioms. The first one that we use in this article is given by:

$$\begin{cases} m_j^i(\{C_i\})(\mathbf{x}) = \alpha_{ij} R_j p(q_j / C_i) / (1 + R_j p(q_j / C_i)) \\ m_j^i(\{C_i\}^c)(\mathbf{x}) = \alpha_{ij} / (1 + R_j p(q_j / C_i)) \\ m_j^i(D)(\mathbf{x}) = 1 - \alpha_{ij} \end{cases} \quad (18)$$

where  $q_i$  is the  $i^{\text{th}}$  classifier (supposed cognitively independent),  $i = \overline{1, N}$ ,  $\alpha_{ij}$  are reliability coefficients on each classifier  $i$  for each class  $j = \overline{1, m}$  (in our application  $\alpha_{ij}=1$ ), and  $R_j = (\max_{q_j, i} (p(q_j / C_i)))^{-1}$ . Hence a mass function is defined for each source and each class. In this approach, the difficulty is the estimation of the probabilities  $p(q_j / C_i)$ . In the case of decision level  $q_j$  is the class given by the classifier  $j$ . Hence the estimation of these probabilities can be made on a learning database by the confusion matrices. In the case of characteristic level, the estimation can be made classically by the frequencies or under assumption of the distribution of these probabilities. For this level the distance approach is easier.

Indeed, in [9] the mass functions are defined by:

$$\begin{cases} m_j^i(\{C_i\} / \mathbf{x}^{(t)})(\mathbf{x}) = \alpha_{ij} \varphi_i(d^{(t)}) \\ m_j^i(D / \mathbf{x}^{(t)})(\mathbf{x}) = 1 - \alpha_{ij} \varphi_i(d^{(t)}) \end{cases} \quad (19)$$

where  $(\mathbf{x}^{(t)})_i$  is a set of learning vector,  $d^{(t)} = d(\mathbf{x}^{(t)}, \mathbf{x})$  is a distance (to be determined) between  $\mathbf{x}$  and  $\mathbf{x}^{(t)}$  and  $C_i$  is the class of  $\mathbf{x}^{(t)}$ .  $\varphi_i$  is a distance function, which verifies:

$$\begin{cases} \varphi_i(0) = 1 \\ \lim_{d \rightarrow +\infty} \varphi_i(d) = 0 \end{cases} \quad (20)$$

Many functions can be used, in [9] Denoeux proposes:

$$\varphi_i(d) = \exp(-\nu_i d^2), \quad (21)$$

where  $\nu_i$  is a positive parameter according to the class  $C_i$ . We will use this function. The distance calculation  $d^{(t)} = d(\mathbf{x}^{(t)}, \mathbf{x})$  can take time if the training database is important. But we can consider only the  $k$  nearest neighbors.

The fundamental difference between the both approaches is that in the first case we have to estimate the probabilities  $p(q_j / C_i)$  and in the

second case the distance  $d$ . For decision level the estimation of  $p(q_j / C_i)$  is very easy, but it is quite difficult to choose an appropriate distance in this case (symbolic distance). On the contrary, for characteristic level, the estimation of  $p(q_j / C_i)$  can be difficult if the distribution is unknown, and Euclidian distance, for example, can be choose for  $d$ . In this paper, we will apply the probabilistic approach for decision level (*i.e.*  $C_{kq}$  outputs of the MLP), and the distance approach for characteristic level (*i.e.*  $o_k$  outputs of the MLP).

Combination of mass functions is based on the orthogonal Dempster-Shafer's rule:

$$m(.) = \bigoplus_{i=\overline{1, N}} \left( \bigoplus_{j=\overline{1, m}} m_j^i(.) \right) \quad (22)$$

In the case of distance approach, this combination is given by:

$$\begin{cases} m_j^i(\{C_i\})(\mathbf{x}) = \frac{1}{L} \left( 1 - \prod_{t \in I_{k,i}} (1 - \alpha_{ij} \varphi_i(d^{(t)})) \right) \\ \prod_{r \neq i} \prod_{t \in I_{k,r}} (1 - \alpha_{rj} \varphi_r(d^{(t)})) \\ m_j^i(D)(\mathbf{x}) = \frac{1}{L} \prod_{r=1}^m \prod_{t \in I_{k,r}} (1 - \alpha_{rj} \varphi_r(d^{(t)})) \end{cases} \quad (23)$$

with  $L$  a normalized constant,  $I_{k,r}$  the set of the neighbors of  $x$  in the class  $C_i$ .

Other conjunctive rules are proposed: the Dempster-Shafer's rule normalized by a conflict measure given by:

$$K = \sum_{B_1 \cap B_2 \cap \dots \cap B_N = \emptyset} \prod_{i=1}^N m_i(B_i) < 1 \quad (24)$$

The Yager's rule [10] redefines  $m(D)$  adding  $K$ . In this article, we choose the Smets's rule [11] that supposes an "open world", for each  $A$  of  $2^D$ :

$$\begin{aligned} m(A) &= \sum_{B_1 \cap B_2 \cap \dots \cap B_N = A \neq \emptyset} \prod_{i=1}^N m_i(B_i) \\ m(\emptyset) &= K \end{aligned} \quad (25)$$

These rules give similar results on our data. The last step of fusion is the decision. In the evidence theory, we can use the maximum of plausibility, maximum of belief or maximum of pignistic probability [11]. We retain the maximum of pignistic probability in this article; the three previous criteria give the same results on our data.

## 5 Database

The database contains 26 sonar images provided by the GESMA (Groupe d'Etudes Sous-Marine de l'Atlantique). These images were obtained with a Klein 5400 sonar with a resolution of 20 until 30 cm in azimuth and 3 cm in range. The sea-bottom deep was between 15 m and 40 m.

These 26 sonar images have been segmented in small-images with a size of 64x384 pixels (*i.e.* of approximately 1152 cmx1152 cm). On Fig. 3 we show a sample of these small-images represented in order to obtain a size of 64x64 pixels.

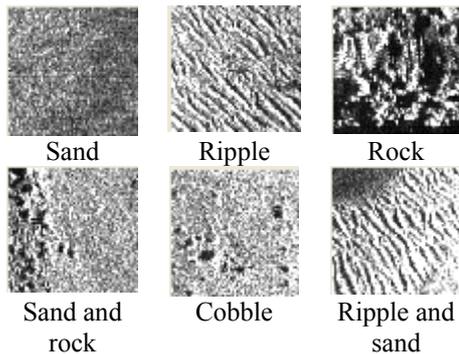


Fig. 3. Sample of small-images with different type of sediment.

Each small-image is characterized manually by the type of sediment (rock, cobbles, sand, ripple, silt) or shadow when the information is unknown (see Tab. 1). Moreover the existence of more than one kind of sediment on the small-image is indicated. In this case the type of sediment affected to the small-image is the most present.

From Tab. 1 we note that the sand sediment is the most represented one. The cobbles sediment is particularly few represented. One of the difficulties of classification step comes from this difference.

Sediment	Effective	%
Rock	915	21.35
Cobbles	33	0.77
Sand	2321	54.62
Ripple	374	8.80
Silt	234	5.50
Shadow	102	2.40
Total	4249	100.00

Tab. 1. Database elements and their effective.

There is 39.7% of small-image with more than one kind of sediment (named patch-worked images).

Note that such database is quite difficult to realize. Indeed, the expert has a subjective experience, and can make a mistake for some small-images.

## 6 Experimental results

The database was randomly divided into three parts. The first one is used for the neural network learning, the second one for the fusion process learning, and the last one for the tests. We repeat this random division 10 times in order to achieve a good estimator of the classification rate, and we analyze the mean percentage of good classification rates defined as the number of good classified small-images on the total of small images (Tab. 2).

cooc	run	wave	Gabor	MLP	Proba	Dist
70.0	50.3	68.9	66.4	50.0	68.8	79.5

Tab. 2. Classification performances ( $\% \pm 2.5\%$ ).

This table shows that the two fusion approaches (Probability based and Distance based) give the best results, and are robust to a bad extraction texture (here run-lengths method), whereas the global multilayer perceptron (MLP) classifying directly all the calculated extraction characteristics is not robust. However the probability based approach performances are not better than co-occurrence matrices or wavelet methods added with a MLP. We can note that the distance based fusion approach gives significantly the best results.

Tab. 3 details the results of this approach.

Rock	87.3
Cobbles	0.9
Sand	84.9
Ripple	61.3
Silt	4.9
Shadow	71.5
No patch-worked	91.3
Patch-worked	63.1

Tab. 3. Detailed classification rates (%).

We note that the best performances are obtained for the rock and sand classes; this is due to the multilayer perceptron learning better for the more numerous type of sediment. The cobbles and silt sediment allow bad results because of the effective low of cobbles and silt images in the database. We notice also that the patch-worked small-images are not good classified (63.1%). This is due to the database constitution.

## 7 Conclusion

We have proposed here a comparison of two classifier fusion approaches applied to a sea-bottom characterization. Both methods are based on the evidence theory. The belief function estimation is based on probability for one fusion approach and on distance for the other one.

The both approaches are robust to bad texture extraction, whereas the used global multilayer perceptron is not. The based distance one allows a significantly improvement. An important problem

for a multilayer perceptron classifier comes from the effective difference of the kind of sediments on our database. The learning for the type of sediment few represented is bad.

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.

## Acknowledgements

The authors thank H el ene Thomas for having provided us some programs and documentations. They want to thank also the GESMA for having provided sonar images.

## References

- [1] H. Thomas, C. Collet, K. Yao et G. Burel, *Some improvements of a rotation invariant autoregressive method. Application to the neural classification of noisy sonar images*, European Signal Processing Conference (Eusipco), Rhodes, Vol. 4, pp 2001-2004, 1998.
- [2] C. Molder, H. Thomas, et A. Quinquis, *Classification des s diments marins par analyse de texture*, RFIA, Angers, France, Vol. 3, pp 799-808, 2002.
- [3] I. Leblond, M. Legris, *Classification et segmentation d'images sonar pour le recalage   long terme*, Journ ee SeaTechWeek, Brest, France, 21-22 october 2004.
- [4] R. Haralick, Statistical and textural approaches to textures, *Proceedings of the IEEE*, Vol. 67(5), pp 786-804, 1979.
- [5] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representations by error propagation, *Parallel Distributed Processing*, vol. 1, 318-362, MIT Press, Cambridge, 1986.
- [6] B. Kosko, *Neural Networks and Fuzzy Systems*, Prentice Hall, Englewood Cliffs, 1992.
- [7] I. Bloch, Some Aspects of Dempster-Shafer Evidence Theory for Classification of multi-modality Medical Images Taking Partial Volume Effect into Account, *Pattern Recognition Letters*, 17: 905-919, 1996.
- [8] A. Appriou, Situation Assessment Based on Spatially Ambiguous Multisensor Measurements, *International Journal of Intelligent Systems*, 16(10): 1135-1166, 2001.
- [9] T. Denoeux, A  $k$ -nearest neighbor classification rule based on Dempster-Shafer Theory, *IEEE Trans. on Systems, Man and Cybernetics*, 25(5): 804-813, 1995.
- [10] R.R. Yager, On the Dempster-Shafer Framework and New Combination Rules, *Information Science*, 41: 93-137, 1987.
- [11] P. Smets, The Combination of the Evidence in the Transferable Belief Model, *IEEE Trans. on*

*Pattern Analysis and Machine Intelligence*, 12(5): 447-458, 1990.