An evidential Markovian model for data fusion and unsupervised image classification

Laurent Fouque, Alain Appriou
DTIM
Onera
29 av de la Division Leclerc 92322 Chatillon-France
fouque@onera.fr

Wojciech Pieczynski
SIM
Institut National des Télécommunications
9 Rue Charles Fourier 91000 Evry-France

Abstract - *In this paper, we deal with the fusion of information and the classification of images supplied by several sensors. By intrinsic characteristics of each sensors, provided informations are usually defined on different set of hypothesis, called frames of discernment. Adapted formalism need to be used to compute the fusion process. We resolve this problem of multi-sensor image fusion and classification in an evidential framework, which is well adapted for the combination of knowledge defined on different frames of discernment. We present two models for merging available informations, a non contextual and a vectorial model which is defined by using a Markov chain structure to represent a* priori *knowledge associated to labelling image. In the Markovian approach, we have that Markovian property is preserved after fusion, which enables us to apply standard classification algorithms. We adopt an unsupervised context in which parameters estimation is done by using a mixture distribution algorithm, the ICE algorithm. We apply these models to satellite images.*

Keywords: Fusion data, Evidential theory, hidden Markov model, image classification

1 Introduction

The current expansion of image acquisition tools has enhanced the development of new theories which allows the best treatment of all available information. In domains as various as satellite and medical image processing, Bayesian theory is generally the most widely used to model information. However, because of its inadequacy for describing imprecise or imperfect data, new more flexible theories such as fuzzy set theory[19], the theory of possibilities[8], and Dempster-Shafer's theory of evidence[17] have been developped.

Our study concerns information fusion and classification of multisource images[5][3][9][14][13]. The aim of a classification process in image processing is to construct a labelling image from multi-sensor images. We are interested in classifying two images supplied by two sensors which have their own physical characteristics. This implies that they do not necessarily detect the same features of the real scene. Based on gray levels intensity of pixels, the interpretation of the observed scene could then be very different from one sensor to another. For example, the first one could detect reflective properties of the objects while the second one could detect thermal characteristics of the objects. Therefore high intensity pixels in one image will correspond to objects with high reflective properties while high intensity for a second image will correspond to warm objects. In that case, it seems natural to consider different referentials $R^{(1)}$ and $R^{(2)}$ associated with each sensor $S^{(1)}$ and $S^{(2)}$ ($R^{(1)}$ adapted to represent reflectivity properties and $R^{(2)}$ adapted to treat heat properties). $R^{(1)}$ could be defined by two classes, the class of strong reflective object and the class of non reflective object, and $R^{(2)}$ by three classes, warm objects, mild object and cold areas. In an operational context, the labelling referential of interest $R^{(3)}$ considered for the classification does not always coincides with $R^{(1)}$ or $R^{(2)}$. As an example, $R^{(3)}$ could correspond two a referential of three classes of interest, the industrial areas, the vegetation and the sea areas. How to link $R^{(3)}$ with $R^{(1)}$ and $R^{(2)}$? Statistical training provides estimates of model parameters on $R^{(1)}$ and $R^{(2)}$ which are natural referential of the images, but no training is available on $R^{(3)}$ which is not directly related to the information provided by the sensors. We define relations between $R^{(1)}$, $R^{(2)}$ and $R^{(3)}$ to use informations on $R^{(1)}$ and $R^{(2)}$ to classify on $R^{(3)}$. In our model, these relations will be modelled by a fuzzy function which will represent the compatibility between classes of $R^{(1)}$ and $R^{(2)}$ with classes of $R^{(3)}$. Few works have been carried out in the litterature to treat the problem of fusion in the context of different referentials associated to sensors[11][6]. In this paper, we develop a markovian model for images classification, adapted to combine these kind of informations.

In the second and third parts, we recall the theoretical probabilistic and evidential notions necessary to the understanding of this article. In the fourth part, we defined a non contextual evidential model for merging informations provided by the sensors. A fusion vectorial model, which is a generalization of the non contextual evidential model, is developped in the fifth part to take into account the spatial correlation between pixels contained in the images. We use a Markov chain model to describe, via a Peano type scan, the non-observable image which represents the real scene. It enables us to significantly decrease the algorithmic complexity of processing encountered when using a Markov fields model. Experimental results are presented in the sixth part. The conclusion gives a summary assessing the methods discussed.

2 Reminders

2.1 The Hilbert scan

A fractal curve first discussed by G. Peano has useful properties in image processing. D. Hilbert published a kind of Peano curve which passes through each grid on a 2-D plane only once. The Hilbert curve is a regular way of transforming multidimensional data arrays into one dimensional ones. The Hilbert scan of image data has recent application applications in image classification[4] and data compression of images[12]. This scan will be used to construct a one-dimensional signal from the two-dimensionals images.

2.2 The Hidden Markov Chains[16]

Let S be the set of pixels corresponding to the images, $card(S) = N$, and $\mathbf{X}$ and $\mathbf{Y}$ two processes, $\mathbf{X}$ represented the labelling image and $\mathbf{Y}$ the observed images. An homogeneous hidden Markov chain is a pair of two processes $(\mathbf{X}, \mathbf{Y}) = ((X_n), (Y_n))_{n \in S}$ with:

- $\mathbf{X}$ is a Markov chain which takes its values in the set of k classes $\Omega = \{w_1, \ldots, w_k\}$.

  The distribution of $\mathbf{X}$ is defined by an initial distribution $\pi_u = P(X_1 = u), u \in \Omega$ and a transition matrix defined by the coefficients $t_{uv} = P(X_{n+1} = v | X_n = u)$. Let $\mathbf{x}$ be a realization of $\mathbf{X}$. The distribution of the process $\mathbf{X}$ is given by:

$$P(X_1 = x_1, \ldots, X_N = x_n) = \pi_{x_1} \times t_{x_1 x_2} \times \ldots \times t_{x_{N-1} x_N} \quad (1)$$

- $\mathbf{Y}$ is a R^d process, where d corresponds to the number of sensors. The conditional density of $\mathbf{Y}$ given $\mathbf{X}$ will be noted $f_{\mathbf{x}}(\mathbf{y})$, $\mathbf{x}$ and $\mathbf{y}$ being a realization of $\mathbf{Y}$.

Consider the three classical hypotheses:
(H1) : the Y_n are independant given $\mathbf{X}$.
(H2) : the distribution of Y_n given $\mathbf{X}$ is equal to its distribution given X_n.
(H3) : sensors are independent given classes.
The joint probability distribution is then given by:

$$\begin{aligned} &P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}) = \\ &\pi_{x_1} \times \prod_{j=1}^{d} f_{x_1}^{(j)}(y_1^{(j)}) \times \\ &\prod_{n=2}^{N} (t_{x_{n-1} x_n} \times \prod_{j=1}^{d} f_{x_n}^{(j)}(y_n^{(j)})) \end{aligned} \quad (2)$$

where $f_{x_n}^{(j)}(y_n^{(j)})$ is the density distribution of $Y_n^{(j)}$ given $X_n = x_n$.

3 The Theory of Evidence - A non contextual approach

The evidence theory developped by Dempster and better formalized by Shafer[17] enables us to represent both uncertainty and imprecision with two functions, plausibility and credibility. These functions are based on the definition of a frame of discernment Ω constituted of k exclusives hypotheses, the k classes. A referential $\Omega^* = 2^{\Omega}$ represents the set of all subsets of Ω. Plausibility and credibility can be expressed with a unique function, the mass function. Mass, plausibility and credibility, which are all defined on Ω^*, characterize the likelihood of any subset of Ω. It can be shown that there exists a bijection between these three functions[10].

3.1 Definitions

The mass function is defined on Ω^* by

$$\left\{ \begin{array}{lcl} m(\emptyset) & = & 0 \\ \sum_{A \subset \Omega} m(A) & = & 1 \end{array} \right. \quad (3)$$

The plausibility and credibility functions are given by:

$$\left\{ \begin{array}{lcl} Pl(\emptyset) & = & 0 \\ Pl(A) & = & \sum_{B \cap A \neq \emptyset} m(B), \forall A \subset \Omega \end{array} \right. \quad (4)$$

$$\left\{ \begin{array}{lcl} Bel(\emptyset) & = & 0 \\ Bel(A) & = & \sum_{B \subset A} m(B), \forall A \subset \Omega \end{array} \right. \quad (5)$$

The focal elements of a mass function are elements of Ω^* such that $m(A) \neq 0$. If the mass function is only defined on the single elements of Ω^*, then mass, plausibility and credibility are equal and correspond to a probability function which we call a Bayesian mass function. A mass function is said to be consonant if its focal elements can be arranged in order so that each focal elements is contained in the following one. Its associated plausibility function is then similar to a possibility function, in the possibility theory.

3.2 Dempster-Shafer's rule of combination

Let $m^{(1)}, \ldots, m^{(d)}$ be the masses associated to d independent information sources defined on the same frame of discernement Ω. It is then possible to combine them according to the Dempster-Shafer's orthogonal combination rule. This rule results in:

$$(m^{(1)} \oplus \ldots \oplus m^{(d)})(A) = \frac{1}{1-K} * \sum_{B_1 \cap \ldots \cap B_n = A} (\prod_{i=1}^{d} m^{(i)}(B_i)) \quad (6)$$

where

$$K = \sum_{B_1 \cap \ldots \cap B_n = \emptyset} (\prod_{i=1}^{d} m^{(i)}(B_i)) \quad (7)$$

is a measure of conflict between the differents sources. This rule has the following property:

the orthogonal combination of any mass sets with a Bayesian mass set is also Bayesian.

After the combination of the different sources of information, each pixel is defined as belonging to a given class acccording to a certain criteria. We choose the maximum plausibility rule which is discussed in [1] to take the decision.

3.3 Refining and coarsening operations

We consider two frame of discernment Θ and Ω. The idea that Ω is obtained from Θ by analysing or splitting some or all the elements of Θ is characterized by the refining operation. A Refining $\mathcal{R}$ is defined by specifying for each classe θ_i of $\Theta = \{\theta_1, \ldots, \theta_k\}$ a subset $\mathcal{R}(\theta_i)$ of $\Omega = \{w_1, \ldots, w_K\}$ consisting of classes into which θ_i has been splitted. The sets $\mathcal{R}(\theta_i)$ should constitute a disjoint partition of Ω. Then a mass function m^1 defined on Θ is transformed to a mass function m^2 on Ω according to the following expression: for all $A \subset \Theta$,

$$m^2(\mathcal{R}(A)) = m^1(A) \quad (8)$$

The coarsening operation, which is the inverse of the refining operation, provides a way to agregate some hypotheses. A mass function m^1 on Θ is obtained from a mass m^2 on Ω according to the following formula: for all $A \subset \Theta$,

$$m^1(A) = \sum_{B \subset \Omega, \hat{\theta}(B) = A} m^2(B) \quad (9)$$

where $\hat{\theta}(B) = \{\theta \in \Theta | \mathcal{R}(\theta) \cap B \neq \emptyset\}$, $B \subset \Omega$.

4 A non contextual fusion model for images classification

In this section, we deal with the problem of merging images provided by two sensors $S^{(1)}$ and $S^{(2)}$, and defined on two different sets of classes $\Omega_1 = \{w_{1_1}, \ldots, w_{1_{k_1}}\}$ and $\Omega_2 = \{w_{2_1}, \ldots, w_{2_{k_2}}\}$. Our goal consists in classifying the images on a third set of classes $\Omega = \{w_1, \ldots, w_k\}$ distinct of Ω_1 and Ω_2. Ω constitutes the set of labels we are interested in for the classification process.

In a non contextual approach, this fusion implies the definition of local mass functions for each pixel n of the images, $m^1_{\Omega_1,n}$ and $m^2_{\Omega_2,n}$, defined on the frames of discernement Ω_1 and Ω_2 respectively. We define these elementary mass functions as follow:
for $1 \leq j \leq 2$,

$$\begin{array}{rlll} m^j_{\Omega_j,n}: & \Omega_j^* & \rightarrow & [0,1] \\ & B & \rightarrow & m^j_{\Omega_j,n}(B) = \oplus_{i=1}^{k_j} m_{ij}(B) \end{array} \quad (10)$$

with

$$\left\{ \begin{array}{lll} m_{ij}(\{w_i\}) & = & 0 \\ m_{ij}(\{w_i\}^C) & = & q_{ij}(1 - C_{ij}) \\ m_{ij}(\Omega_j) & = & 1 - q_{ij}(1 - C_{ij}) \end{array} \right. \quad (11)$$

These masses characterizing the likelihood of each classes w_i. C_{ij} corresponds to a quantity capable of discriminating the class w_i according to the observations. Severals forms of C_{ij} could be used, each adapted to different contexts[2]. In this paper, we define them in terms of the conditional density distribution of the observations Y_n given the classes $X_n = w_i$, $P(Y_n^j = y_n^j | X_n = w_i)$. A quality factor q_{ij}, $0 \leq q_{ij} \leq 1$, is combined with each likelihood C_{ij}. To take the heterogeneity of the frames of discernment into account, we consider for each sensor j, $1 \leq j \leq 2$, two fuzzy membership functions S^1 and S^2:
for $1 \leq j \leq 2$,

$$\begin{array}{rlll} S^j: & \Omega_j \times \Omega & \rightarrow & [0,1] \\ & (u,v) & \rightarrow & S^j(u,v) \end{array} \quad (12)$$

These similarity functions, defined on the cartesian product of the frames of discernment, correspond to the degree of compatibility between classes of Ω^j with classes of Ω. Their determination depends on the context. In a supervised context, these degrees of compatibility will be supplied by an expert who possesses an *a priori* knowledge concerning these compatibility links. In an unsupervised one, they should be estimated from the images. To integrate these informations in our model, we construct two possibility measures by considering α-cuts of the fuzzy functions S^j, $1 \leq j \leq 2$, these α-cuts defining sets that includes previous ones. From these possibility measures, we elaborate two consonant mass functions $\eta^j_{\Omega_j \times \Omega}$, based on the cartesian product $\Omega_j \times \Omega$.

At this step , we have for each sensors $(1 \leq j \leq 2)$ two kinds of informations:

- an information provided by the sensor and modelized by a mass function $m^j_{\Omega_j,n}$: $2^{\Omega^j} \rightarrow [0,1]$
- an *a priori* knowledge describing the compatibility between classes of Ω^j and Ω and modelized by a consonant mass function $\eta^j_{\Omega_j \times \Omega}$: $2^{\Omega^j \times \Omega} \rightarrow [0,1]$

To combine $m^j_{\Omega_j,n}$ and $\eta^j_{\Omega_j \times \Omega}$ according to the dempster-Shafer's rule of combination, we need to define them on a commun referential. By refining, we construct from the mass function $m^j_{\Omega_j,n}$ two new mass functions $\hat{m}^j_{\Omega_j\times\Omega,n}$ on $2^{\Omega^j\times\Omega}$. The combination is then feasible and correspond to the following expression: for $1 \leq j \leq 2$,

$$M^j_{\Omega_j\times\Omega,n}(.) = (\hat{m}^j_{\Omega_j\times\Omega,n} \oplus \eta^j_{\Omega_j\times\Omega})(.) \qquad (13)$$

These mass functions characterize the informations associated with each sensor. To perform the classification process on the referential of interest Ω, it is necessary to construct two new mass functions $\hat{M}^j_{\Omega,n}$ deduced from $M^j_{\Omega_j\times\Omega,n}$ and defined on Ω. We elaborate them by using the coarsening operator, which provides mass functions on the referential of interest:

$$\hat{M}_{\Omega,n} : \; 2^{\Omega} \rightarrow [0,1] \qquad (14)$$

These functions could then be combined according to the Dempster-Shafer's rule of combination:

$$\mathcal{M}_{\Omega,n}(.) \;=\; (\hat{M}^1_{\Omega,n} \oplus \hat{M}^2_{\Omega,n})(.) \qquad (15)$$

The resulting mass function represents all the information available, including sensors and similarity classes informations. Define on the referential Ω, it could be used to perform the classification process.
Added to the information, we introduce an *a priori* information associated with the labelling image and represented by a Bayesian mass function $M^0_{\Omega,n}$ defined on the referential Ω by:

$$M^0_{\Omega,n}(u) \;=\; \pi_u \quad , u \in \Omega \qquad (16)$$

where $\pi_u, u \in \Omega$ are probabilities to observe the classes u in the real labelling image.

Decision rule:
To perform the classification process and attribute a label to each pixels, we use the maximum of plausibility[1], which is in a Bayesain context, equal to the maximum of mass. This rule is given after combination of $M^{(0)}_{\Omega,n}$ with $\mathcal{M}_{\Omega,n}$ by the relation:

$$\widehat{x_n} = arg \max_{u\in\Omega} M^0_{\Omega,n}(u) \sum_{u \subset B \neq \emptyset} \mathcal{M}_{\Omega,n}(B) \qquad (17)$$

5 A vectorial fusion model for images classification

In remote sensing, available information have a spatial structure inherent to the real scene. To take this constraint into account, it is neccessary to elaborate fusion processes which integrate this contextual notion. In this section, we then develop a markovian vectorial model, which is a generalization of the non contextual model, to represent spatial correlation between the pixels of the labels images. This model implies the definition of vectorial mass and plausibility functions.

5.1 Definitions

5.1.1 Mass function

Let $\Theta = \Omega^N$, the cartesian product of the frame of discernment Ω. A vectorial mass function m defined on Ω^N will be given by:

$$m_\Theta : \begin{array}{ccc} \Theta^* & \rightarrow & [0,1] \\ \hat{A} & \rightarrow & m_\Theta(\hat{A}) \end{array} \qquad (18)$$

To represent the images, the set of all acceptable global configurations are elements of $(\Omega^* \setminus \emptyset)^N$. Indeed, each pixel $X_n, 1 \leq n \leq N$ could only take value in $\Omega^* \setminus \emptyset$. Then we define a set of evidence ν on $(\Omega^* \setminus \emptyset)^N$ which represents the evidence distributing to each acceptable global configuration. We constrain ν to be normalized, that is to say $\sum_{X\in\Omega^*\setminus\emptyset} \nu(X) = 1$. We notice that ν is not a mass function since it is not define on the set of subsets of Ω^N. However we have the following proposition:

Proposition 1:
There exist a one to one function h between the acceptable global configurations $(\Omega^* \setminus \emptyset)^N$ and a subset T of $(\Omega^N)^*$.

For exemple, the injective function

$$h : \begin{array}{ccc} (\Omega^* \setminus \emptyset)^N & \rightarrow & (\Omega^N)^* \\ (A_1, \ldots, A_N) & \rightarrow & A_1 \times \ldots \times A_N \end{array} \qquad (19)$$

induces a bijection from $(\Omega^* \setminus \emptyset)^N$ into $h((\Omega^* \setminus \emptyset)^N) \subset (\Omega^N)^*$.

According to this proposition, we can define the mass function m_Θ on Θ^* in terms of ν.
Let the vectorial function m_Θ be:

$$\left\{ \begin{array}{lcll} m_\Theta(\hat{A}) & = & \nu(A), & \hat{A} = h(A) \\ m_\Theta(\hat{A}) & = & 0, & otherwise \end{array} \right. \qquad (20)$$

We verify that m_Θ is a mass function since it satisfies (18).

5.1.2 Plausibility function

We define the plausibility function by:

$$Pl \;:\; \begin{array}{ccccc} (\Omega^N)^* & \rightarrow & [0,1] & & \\ \hat{A} & \rightarrow & Pl(\hat{A}) & = & \sum_{\hat{B}\cap\hat{A}\neq\emptyset} m_{\Omega^N}(\hat{B}) \end{array} \qquad (21)$$

which is equivalent, for the mass function we consider, to $Pl_\Theta : (\Omega^N)^* \rightarrow [0,1]$ with for $\hat{A} \in (\Omega^N)^*$:

$$\begin{array}{ll} Pl_\Theta(\hat{A}) = \sum_{B\in(\Omega^*\setminus\emptyset)^N, B\cap A\neq\emptyset} \nu(B), & \hat{A} = h(A) \\ Pl_\Theta(\hat{A}) = \; 0 & otherwise \end{array}$$

5.2 vectorial fusion model

Consider a global mass functions m_{Ω^N} to represent the informations contained in the images and defined on the frame of discernement Ω. We decide to use an assumption of independence concerning the pixels of the observed images, even if that corresponds to a strong assumption. In this context, we define m_{Ω^N} by:

$$\begin{array}{lll} m_{\Omega^N}: & (\Omega^N)^* & \rightarrow [0,1] \\ & \hat{B} & \rightarrow m_{\Omega^N}(\hat{B}) = \nu(B), \\ & & \hat{B} = h(B), B \in (\Omega^*)^N; \\ & & m_{\Omega^N}(\hat{B}) = 0, \\ & & otherwise \end{array} \quad (22)$$

and

$$\nu(B) = \prod_{n=1}^{N} \mathcal{M}_{\Omega,n}(B_n)$$

Added to the information provided by the sensors, we introduce an a *priori* global information represented by a vectorial Bayesian mass function $M^{(0)}_{\Omega^N}$ defined on the referential Ω^N. This mass function is constructed by introducing a Markov chain structure as follows :

$$\begin{array}{lll} M^0_{\Omega^N}: & (\Omega^N)^* & \rightarrow [0,1] \\ & \hat{B} & \rightarrow M^0_{\Omega^N}(\hat{B}) = \nu^0(u), \\ & & \hat{B} = h(u), u \in (\Omega)^N; \\ & & M^0_{\Omega^N}(\hat{B}) = 0, \\ & & otherwise \end{array} \quad (23)$$

and

$$\nu^0(u) = \pi_{u_1} \times t_{u_1 u_2} \times \ldots \times t_{u_{N-1} u_N}$$

π_u corresponds to the probability of observing the class $u \in \Omega$ in the real labelling image and t corresponds to the transition matrix of the Markov chain.

Proposition 2:
The Dempster-Shafer's combination rule applied to m_{Ω^N} and $M^0_{\Omega^N}$ results in an a posteriori distribution $P(\mathcal{X}|\mathcal{Y})$ of a classical hidden Markov chain $(\mathcal{X},\mathcal{Y})$ defined by:

$$P(\mathcal{X} = u) = \pi_{u_1} \times t_{u_1 u_2} \times \ldots \times t_{u_{N-1} u_N}$$

and

$$P(\mathcal{Y}_n|\mathcal{X}_n = A_n) = g(A_n)$$

where

$$g(A_n) = \sum_{A_n \subset C_n, C_n \in \Omega^*} \mathcal{M}_{\Omega,n}(C_n)$$

This proposition provides a way to compute the classification process. Using the maximum of plausibility, we classify each pixels according to the following decison rule:

$$\widehat{(x_1,\ldots,x_N)} = arg \max_{(A_1,\ldots,A_n)\in\Omega^N} \pi_{A_1} g(A_1) \times \prod_{n=2}^{N} t_{A_{n-1}A_n} g(A_n) \quad (24)$$

In that case, the classification obtained by maximising the plausibility is similar to the MAP criteria applied to the hidden Markov chain $(\mathcal{X},\mathcal{Y})$. Solution is computed by the viterbi algoritm. It is also possible to consider the MPM criteria computed by the calculation of "Forward-Backward" probability.

6 Parameter estimation algorithm

Parameters are usually unknown and we need to estimate them from the observations. In the following, we work in an unsupervised context. Different estimation algorithms are available. The iterative EM algoritm (Expectation-Maximisation) [7] maximizes the likelihood of the observations. Modifications of this algorithm have been proposed to improve convergence of the algorithm to a global maximum of the likelihood [15]. In this paper, we use a recent algorithm for mixture distribution estimation, presented in [4]. We apply this algorithm to the case of our model.

7 Application

7.1 Fusion of optic and radar images

To evaluate the quality of the model proposed, we carried out segmentations of synthetic and real images.

7.2 Fusion of synthetic images

Let us consiser as the real scene, the image presented in figure (1). We construct two images, which are considered as two "pseudo" real scenes, representing the classes detected by the sensors according to their own physical characteristics and compatible with the real scene (figure (2)). We corrupte these two images with normal noises (for each class: $\mathcal{N}(0,1)$ and $\mathcal{N}(2,1)$ for the first image, $\mathcal{N}(0,1)$, $\mathcal{N}(1,1)$ and $\mathcal{N}(2,1)$ for the second image). The segmentation problem is resolved by using the markovian model presented previously.
An a *priori* information on the labels image is defined by a Bayesian mass set M^0_n defined on a referential $\Omega = \{w_1, w_2\}$ by

$$m^0_{\Omega,n}(u) = \pi_u \quad , u \in \Omega \quad (25)$$

For the two sensors, we consider the mass functions: for $1 \le j \le 2$,

$$\begin{array}{lll} m^j_{\Omega_j,n}: & \Omega^*_j & \rightarrow [0,1] \\ & B & \rightarrow m^j_{\Omega_j,n}(B) = \oplus_{i=1}^{k_j} m_{ij}(B) \end{array} \quad (26)$$

with

$$\left\{\begin{array}{lcl} m_{ij}(\{w_i\}) & = & 0 \\ m_{ij}(\{w_i\}^C) & = & (1-P(Y_n^j = y_n^j | X_n = w_i)) \\ m_{ij}(\Omega_j) & = & 1-(1-P(Y_n^j = y_n^j | X_n = w_i)) \end{array}\right. \quad (27)$$

We introduce fuzzy functions to model similarities between classes of the different referentials, here Ω_1 and Ω_2 with Ω. Figure (4) provides results of the segmentation. We notice that fusion allows to classify quite well (the object on the right of left image (figure(3)) is well classify; the object in the center of the right image (figure(3)) has disappeared). 4.2% of pixels are misclassified.

7.3 Fusion of optic and radar images

We are now interested in classifying a scene of "Istres", an area in the south of france. We expect to segment in three classes ($\Omega = \{city, vegetation, water\}$). We use two complementary images, a radar and an optic images, presented in figure (5). An evidential approach is well adapted to treat these heterogeneous informations.

7.3.1 Classification in a non Contextual Approach

An *a priori* information on the labels image is defined by a Bayesian mass set M_n^0 defined on the referential Ω by

$$m^0_{\Omega,n}(u) \quad = \quad \pi_u \quad , u \in \Omega \qquad (28)$$

For ERS and LANDSAT sensors, we use the previous model to construct the mass functions $m^{(1)}_{\Omega_1,n}$ and $m^{(2)}_{\Omega_2,n}$, defined respectively on the referential $\Omega_1 = \{Strong\_echo,\ medium\_echo,\ low\_echo\}$ and $\Omega_2 = \{very\_high\_density,\ high\_density,\ low\_density,\ very\_density\}$:
for $1 \leq j \leq 2$,

$$\begin{array}{rcl} m^j_{\Omega_j,n}: \ \Omega_j^* & \rightarrow & [0,1] \\ B & \rightarrow & m^j_{\Omega_j,n}(B) = \oplus_{i=1}^{k_j} m_{ij}(B) \end{array} \qquad (29)$$

with

$$\left\{\begin{array}{lcl} m_{ij}(\{w_i\}) & = & 0 \\ m_{ij}(\{w_i\}^C) & = & (1-P(Y_n^j = y_n^j | X_n = w_i)) \\ m_{ij}(\Omega_j) & = & 1-(1-P(Y_n^j = y_n^j | X_n = w_i)) \end{array}\right. \quad (30)$$

Theses mass functions model the evidence we have for pixels to belong to each class according to the measures given by the sensors. We consider two fuzzy functions to model similarities between classes of the different referentials and construct two mass functions by $\alpha-$cuts.
The fusion of these informations requires to be defined on the same referential. According to the processes presented in the paragraph (4), we redefine $m^1_{\Omega_1,n}$ and $m^2_{\Omega_2,n}$ on Ω by refining and coarsening. We then combine with $m^0_{\Omega,n}$ according to dempster-Shafer's rule of combination.

7.3.2 Classification in a Markovian context

To take the spatial structure of the labels image into account, we consider an *a priori* information defined by a Bayesian mass function $M^{(0)}_{\Omega^N}$ on the referential Ω^N. This mass is characterizes by a markov chain model:

$$\begin{array}{rcl} M^0_{\Omega^N}(u_1,\ldots,u_N) & = & P(X_1 = u_1,\ldots,X_N = u_N) \\ & = & \pi_{u_1} a_{u_1 u_2} \times \ldots \times a_{u_{N-1} u_N} \end{array}$$

Information supplied by ERS and SPOT sensors defined by:

$$m_{\Omega^N}(A_1,\ldots,A_N) \quad = \quad \prod_{n=1}^N m_{\Omega,n}(A_n)$$

According to the processes presented in the paragraph (5), and after the combination of $m^{(0)}_{\Omega^N}$ and m_{Ω^N} with Dempster-Shafer's rule, we obtain results of classification presented in figure (5) (We recall that the combination provides the distribution of a hidden Markov chain).
We observe that both ERS and LANDSAT monosensor segmentation could not possess power enough to discriminate classes considered. ERS image do not dinstinguish well between city and relief (both strong echo), and LANDSAT image mistake vegetation, culture and part of water. The evidential model, which uses these complementary informations, discriminates better the classes of interest. Water and vegetation are well detected by the fusion, and only very high density culture have been wrongly detected and confused with city class (right down in th image). A new information, as another LANDSAT canal could eventually improve the discrimination power of the model and allow to differenciate between high density culture and city class.

8 Conclusion

As the information provided by a single sensor could be incomplete or imprecise, it is of interest to merge this information with a new one provided by another sensor to obtain a better description of the unobservable scene. This work concerns the classification of multisensor images defined on different set of classes. We present a evidential model to take into account the complementarity of images. We introduce a Markovian structure to model spatial interaction of data. Experimental results show the advantages of the Makovian evidential modeling compares to monosensor model.

References

[1] Appriou, A. "Probabilities and uncertainty in multi-sensor Data Fusion". Revue Scientifique et Technique de la Défense, n. 11, pp 27-40, 1991.

[2] Appriou, A. "Multisensor signal processing in the framework of the theory of evidence".

NATO/RTO - Lecture Series 216 on “Application of Mathematical Signal Processing Techniques to Mission Systems”, November, 1999

[3] Bendjebbour, A. and Pieczynski, W. "Segmentation d'images multisenseur par fusion evidentielle dans un contexte markovien". Traitement du Signal, vol 19, n. 5, pp 465-475, 1997.

[4] Benmiloud, B. and Pieczynski, W. "Estimation des paramètres dans les chaînes de Markovs cachées et segmentation d'images". Traitement du Signal, vol 12, n 5, pp 433-454, 1995.

[5] Bloch,I "some aspect od Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account". Pattern Recognition Letters, vol 17, n. 8, pp 905-916, 1996.

[6] Chatalic, P. "Raisonnement déductif en présence de connaissances imprécises et incertaines: Un système basé sur la théorie de Dempster-Shafer". PhD, Université Paul Sabatier de Toulouse, 1986.

[7] Dempster, A.P., Laird, N. M. and Rubin, D. B. "Maximum likelihood from incomplete data via the EM algorithm". Journal of the Royal Statistical Society, B, vol 39, pp 1-38, 1977.

[8] Dubois, D. and Prade, H. "Théorie des possibilités : Applications à la representation des connaissances en informatique". Masson, 1988.

[9] Giordana, N. and Pieczynski, W. "Estimation of generalized multisensor hidden markov chains and unsupervised imge segmentation". IEEE Trans. on PAMI, vol 19, n. 5, pp 465-475, 1997.

[10] Guan, J. and Bell, D.A "Evidence theory and its applications". North-Holland Amsterdam, 1991.

[11] Janez, F. and Appriou A. "Theory of evidence and non-exhaustive frames of discernment". Traitement du Signal, vol 13, n. 3, pp 237-250, 1996.

[12] Kamata, S."Data compression for gray images using hilbert scan", System and Computer in Japan, vol 28, n. 9, pp 26-34, 1997.

[13] Lehegarat-Mascle, S. and Bloch, I. and Vidal-Madjar, D. "Application of Dempster-Shafer Evidence. Theory to unsupervised classification in multisource remote sensing". IEEE trans. on Geoscience and Remote Sensing, vol 35, n. 4, pp 1018-1031, 1997.

[14] Lehegarat-Mascle, S. and Bloch, I. and Vidal-Madjar, D. "Introduction of Neighborhood Information in Evidence Theory and Application to Data Fusion of Radar and Optical Images with Partial Cloud Cover". Pattern Recognition letters, vol , n. 31 , pp -, 1998.

[15] McLachlan, G.J. and Krishnan, T. "The EM algorithm and Extensions" Wiley Series in Probability and Statistics, 1997.

[16] Rabiner, L.R. "A tutorial on hidden markov models and selected applications in speech recognition". Proceeding of th IEEE, vol 77, n. 2, pp 257-286, 1989.

[17] Shafer, G. "A mathematical theory of evidence". Princeton university Press, 1976.

[18] Smets, P. "About updating". Uncertainty in Artificial Intelligence, vol 6, n. 5, pp 378-385, 1991.

[19] Zadeh, L.A. "Fuzzy sets as a basis for theory of possibility". Fuzzy Sets and Systems 1, pp 3-28, 1978.

Figure 1: real scene

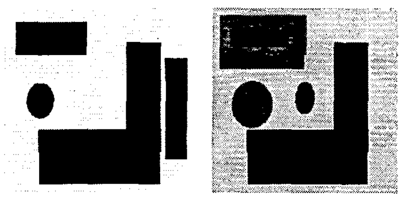


Figure 2: pseudo real scene

Figure 3: noisy images

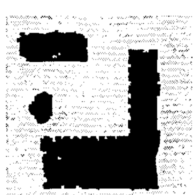
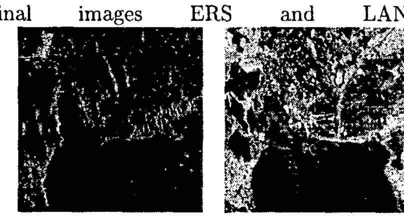
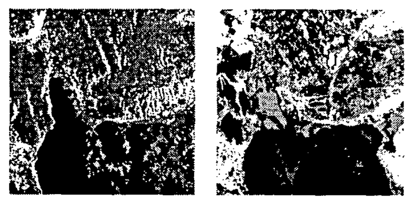


Figure 4: Sementation result

Original images ERS and LANDSAT





Segmentation from ERS only; LANDSAT only

Figure 5: segmentation result using ERS and LANDSAT with the evidential Markovian model