First International Conference on Multisource-Multisensor Information Fusion, July 6 - 9, 1998, Las Vegas, Nevada, USA, pp. 595-600.

Unsupervised Image Segmentation Using Dempster-Shafer Fusion in a Markov Fields Context

A. Bendjebbour
Laboratoire de Statistique Théorique et Appliquée
Université Paris VI
4, Place Jussieu, 75005 Paris, France.

W. Pieczynski
Département Signal et Image
Institut National des Télécommunications
9, rue Charles Fourier, 91000 Evry, France.

Abstract - The Dempster-Shafer combination rule can be of great utility in multisensor image segmentation. In addition, the approach based on theory of evidence can be seen as generalizations of the classical Bayesian approach, which is often used in the Hidden Markov Field Model context. Finally, some recent works allow one to use the Dempster-Shafer combination rule in the Markovian context, and different methods so obtained can greatly improve the effectiveness of Markovian methods working alone. The aim of this paper is to make these methods unsupervised by proposing some parameter estimation algorithms. In order to do so, we use some recent methods of generalized mixture estimation, which allows one to estimate mixtures in which the exact nature of components is not known.

Key Words: Dempste-Shafer fusion, Markov fields, unsupervised image segmentation, parameter estimation, multisensor data.

1. Introduction

Fusion of information provided by numerous sensors, which possibly are of different nature, has met with increased interest in different fields of signal and image processing, particularly, in satellite and medical imaging. One can deal with this problem by the use of probabilistic models and statistical processing. In particular, it is well known that in a Bayesian context the Hidden Markov Field model based segmentation methods may be of exceptional efficiency [2, 5, 6, 8, 13, 14, 15, 20]. Otherwise, fusing information supplied by different sensors can be performed by exploiting the theory of evidence [1, 11, 12, 16, 17, 18]. A piece of information is attached to each sensor via a "fuzzy" measure, which gives, in a particular case, a classical probability measure. Then the fusion is performed by the so-called Dempster-Shafer combination rule. We notethat when at least one sensor provides a classical probabilistic measure, the fusion result is a classical probabilistic measure. Thus, in the image segmentation context of interest, when at least one sensor gives a classical probability measure we can use, after fusion, classical Bayesian decision rules to perform segmentation. In some situations, and in particular when the knowledge of the probability distribution of some sensors is not precise enough, replacing these probability distributions with fuzzy measures can improve the final segmentation, based on the fusioned probability measure. These possible advantages of the Dempster-Shafer fusion can be exploited in the hidden Markov fields context, which allows one to merge the advantages of both models. We proposed in [3] some heuristic manners of Dempster-Shafer fusion which take into account the Markovian structure, and numerous simulations show that merging the two approaches can be of interest. The latter study led us to propose "evidential" hidden Markov models, which can be hidden Markov fields [4] or hidden Markov chains [10]. The aim of this paper is to study how the evidential hidden Markov field based image segmentation, which is in this context a Bayesian classification, can be rendered unsupervised. More precisely, the problem is to estimate different parameters from the observations alone. In the classical hidden Markov field case, different solutions to the difficult parameter estimation problem have been proposed [2, 5, 13, 15, 20] and, although it is very difficult to advance any theoretical results, the methods generally perform well. Here we propose an original method, inspired by the "generalized" mixture estimation methods proposed in [7, 9], to solve this problem.

The organization of the paper is as follows. In the next section we briefly recall the classical multisensor hidden Markov field model, and section 3 is devoted to a brief description of the "fuzzy", or "evidential" measures and Dempster-Shafer combination rule. The Dempster-Shafer fusion in a Markovian context is specified in section 4, and the parameter estimation method we propose is described in section 5, and section 6 contains some simulation results. Conclusions and perspectives for further work are presented in section 7.

2. Multisensor Hidden Markov Fields

Given the set S of pixel, we consider two sets of random variables $X=(X_s)_{s\in S}$, $Y=(Y_s)_{s\in S}$ called "random fields". The field X models the unobservable class field (each X_s takes its values in a finite set of classes $\Omega=\{\omega_1,\ldots,\omega_k\}$), and Y models the observations (for m sensors, each $Y_s=(Y_s^1,\ldots,Y_s^m)$ takes its values in R^m). The segmentation problem consists in estimating the unobserved realization $X=x$ of the field X from the observed realization $Y=y$ of the field Y, where $y=(y_s)_{s\in S}$ are m digital images. The field $X=(X_s)_{s\in S}$ is said to be Markovian with respect to a neighborhood V if its distribution can be written as

$$P_X[x]=\gamma e^{-\sum_{e\in E}\Psi_e(x_e)} \qquad (2.1)$$

where $U(x)=\sum_{e\in E}\Psi_e(x_e)$ is called the energy. E is the set of cliques (a clique being a subset of S which is either a singleton or a set containing mutual neighbours with respect to V), x_e is the restriction of x to e, and Ψ_e is a function, which depends on e only, and which takes its values in R. Assuming that

(i) the random variables (Y_s) are independent conditionally to X;

(ii) the distribution of each Y_s conditional to X is its distribution conditional to X_s;

(iii) the random variables $Y_s^1,\ldots,Y_s^m$ are independent conditionally to X_s,

one can show that all the distributions of Y conditional to X are defined, for k classes, and m sensors, by km distributions on R. To be more precise, let f_i denote the density on R^m of the distribution of Y_s conditional to $X_s=\omega_i$ (f_i is a product of m densities $f_i^1,\ldots,f_i^m$ on R). Thus the distribution of (X,Y) is defined by the functions Ψ_e and the densities f_i. It is then possible to perform the segmentation by the maximum posterior mode (MPM) method: $\hat{x}=\mathrm{MPM}(y)$ if for each pixel $s\in S$

$$\hat{x}_s=\arg\max_{x\in\Omega} P[X_s=x_s|Y=y] \qquad (2.2)$$

by using the algorithm of Marroquin *et al.* [14].

3. Evidential measures and Dempster-Shafer combination rule.

There are three equivalent ways to introduce the evidential measures on $\Omega = \{\omega_1, ..., \omega_k\}$: plausibilities, belief functions or mass functions. In this work we will adopt the representation by mass functions. Let us denote by $\Omega^* = \{\Omega_1, ..., \Omega_{2^k}\}$ the set of subsets of Ω. A mass function M is a probability on Ω^* verifying $M[\emptyset] = 0$. Let us consider m mass functions $M_1^s, ..., M_m^s$. Roughly speaking, $M_1^s, ..., M_m^s$ will model the information contained in the observation of m sensors at pixel s. The Dempster-Shafer combination rule, which enables one to aggregate these different pieces of information, is as follows:

$$M^s[A] = c \sum_{A_0 \cap ... \cap A_m = A \neq \emptyset} \left[\prod_{j=0}^{m} M_j^s[A_j] \right] \quad (3.1)$$

with c the normalising constant. The probability M^s is generally denoted by $M^s = M_1^s \otimes ... \otimes M_m^s$. The mass functions can be seen as generalizations of probability distributions in the following way: when the mass of every set but the singletons is null, it can be assimilated to a probability distribution. Such a mass is called "probabilistic", or "Bayesian". An important property is that if at least one mass function among $M_1^s, ..., M_m^s$ is probabilistic, then $M^s = M_1^s \otimes ... \otimes M_m^s$ is also probabilistic.

4. Dempster-Shafer fusion in a Markovian context

The classical Markov field, whose distribution is given by (2.1), can be extended to an "evidential" Markov field, whose distribution is given on $(\Omega^*)^N$, where N is the set of pixels, by an analogous formula:

$$M^0[x^*] = \gamma * e^{-\sum_{e \in E} \Psi_e^*(x_e^*)} \quad (4.1)$$

We can thus consider that prior information on x^* is given by (4.1). Otherwise, the information contained in the observation $Y_s^j = y_s^j$ at a given pixel s is modelled by a mass function M_j^s defined on Ω^*. More precisely, for a given observation $y_s = (y_s^1, ..., y_s^m)$, the mass functions $M_1^s, ..., M_m^s$ are defined by

$$M_j^s(A) = \frac{g_A^j(y_s^j)}{\sum_{B \in \Omega^*} g_B^j(y_s^j)} \quad (4.2)$$

in which g_A^j are probability densities on R. Recall that $g_A^1, ..., g_A^m$ correspond to $f_i^1, ..., f_i^m$ of section 2.

Let us consider $M^s = M_1^s \otimes ... \otimes M_m^s$. Then, M^0 models the prior information, $M^{obs} = (M^s)_{s \in S}$ models the information contained in the observations, and $M = M^0 \otimes M^{obs}$ models the whole information available about x^*. Of course, the direct application of (3.1) to calculate M is not feasible in a general case and some complementary hypotheses are needed. We have shown in [4] that if (i), (ii), and (iii) section 2 are verified and if either M^0 or $(M^s)_{s \in S}$ is probabilistic, that the fusion is feasible and the result is a probabilistic posterior distribution of a classical hidden Markov field. Thus the classical MPM segmentation is feasible. Let us note that this holds when one at least of the mass functions $M_1^s, ..., M_m^s$ is probabilistic.

5. Parameter estimation

Although the considerations to follow are quite general, in order to simplify things, we shall place ourselves in a particular case. Let us consider two sensors Y^1, Y^2 and $\Omega = \{a, b, c\}$. The first sensor is probabilistic and the second one is evidential, with the corresponding mass function defined on $\Omega^* = \{A, B, C, D\}$, with $A = \{a\}$, $B = \{b\}$, $C = \{c\}$, and $D = \Omega = \{a, b, c\}$. Thus we have to determine g_a, g_b, and g_c concerning the first sensor, and g_A, g_B, g_C, and g_D concerning the second one. Estimating these functions from the observations is the mixture estimation problem and we propose to solve it by applying a new variant of a recent method of "generalized" mixture estimation. A mixture is called generalized when the form of each component is not known; however, it belongs to a given set of forms. For instance, if each of the densities g_a, g_b, and g_c can be Normal or exponential, we have 8 possibilities of classical mixtures and the additional difficulty is to determine in which case we lie.

We consider the Ising model, whose distribution is defined by $\alpha \in R$ for X, and propose the following method:

1) Consider (X, Y^1), which is a classical hidden Markov field described in section 2. We assume that each of the densities g_a, g_b, and g_c can be a Normal, Beta, or Gamma density. We apply our method, whose novelty is specified below, to estimate this Markovian generalized mixture.

2) Consider (Y_s^2) without any Markovian structure. The distribution of Y_s^2 is thus a classical mixture on R of four distributions g_A, g_B, g_C, and g_D. We could still apply our method to treat this mixture as a generalised mixture, although in simulations below, we consider that it is a Gaussian mixture and we estimate it with the classical ICE [15].

The general mixture estimation method proposed in [9], called the ICE-GEMI algorithm, is an iterative method: at step q, let α^q and g_a^q, g_b^q, g_c^q be current prior parameters and current densities g_a, g_b, and g_c. The updating is as follows:

(a) Simulate x^q, a realization of X, according to its α^q and g_a^q, g_b^q, g_c^q based distribution conditional to $Y^1 = y^1$. Calculate $\alpha^{q+1} = \hat{\alpha}(x^q)$, with $\hat{\alpha}$ the Younes method [19].

(b) For $i = a, b, c$, consider $S_i^q = \{s \in S / x_s^q = i\}$. Let $y_i^q = (y_s)_{s \in S_i^q}$. For each $i = a, b, c$ estimate, from y_i^q, the three "candidate" densities: f_i^1 (Normal), f_i^2 (Beta), and f_i^3 (Gamma).

(c) For $i = a, b, c$, choose between f_i^1, f_i^2, and f_i^3 using a decision rule D, which gives $D(y_i^q) \in \left\{ f_i^1, f_i^2, f_i^3 \right\}$.

(d) Update g_a, g_b, g_c by putting $(g_a^{q+1}, g_b^{q+1}, g_c^{q+1}) = (D(y_a^q), D(y_b^q), D(y_c^q))$.

The novelty of our method is situated at the decision rule D level. In [7] the rule D is based on the use of the Pearson system in which one calculates the skewness and the kurtosis, and in [9] the rule D is the minimization of the Kolmogorov distance. The decision rule we propose is based on kernel estimation; the step (c) becomes:

(i) For $i = a, b, c$, calculate $\hat{f}_i(y) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{y - y_j}{h_n}\right)$, where (y_j) are in y_i^q and K is the Normal kernel: $K(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$;

(ii) take among f_i^1, f_i^2, and f_i^3 the density f_i^r such that $\left\| f_i^r - \hat{f}_i \right\|_\infty = \min_{1 \le t \le 3} \left\| f_i^t - \hat{f}_i \right\|_\infty$.

This procedure turns out to perform better, at least in the setting of our experiments, than the Pearson system based method.

6. Experiments

As mentioned above, we consider here g_A, g_B, g_C, and g_D Normal, with variance 1 and means 0, 2, 4, and 1, respectively, whose parameters are estimated by classical ICE. Thus, concerning these densities, we consider a particular case, though a generalized ICE could easily be applied here. The forms and parameters of g_a, g_b, g_c are given in Table 1, which also contains the Bayesian error ratio (performed from the Bayesian sensor Y^1 only), and the "fusioned" error ratio (performed from the both Y^1 and Y^2 after their fusion).

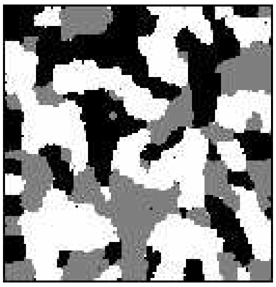
Table 1

B, G, N: Beta, Gamma, Normal distributions, respectively, with the corresponding parameters. MPM (Y^1): Error ratio obtained by the MPM method using the only probabilistic sensor Y^1. Fusion MPM: Error ratio obtained by the MPM method after the Dempster-Shafer fusion of the sensors Y^1 and Y^2. ICE The classical ICE assuming all distributions normal.

Case	Class	*a*	*b*	*c*	MPM (Y^1)	Fusion MPM
1	True Laws	B(7.0, 7.0, 0.0, 20.0)	G(1.5, 2.0, 5.0)	N(2.0, 1.0)	3.63	1.86
2	ICE	N(9.8, 5.4)	N(7.8, 5.6)	N(1.5, 1.0)	4.54	2.22
3	ICE-PEAR	N(8.7, 5.8)	G(1.2, 2.6, 6.0)	N(2.4, 1.3)	9.93	5.69
4	ICE-KERNEL	B(4.3, 6.0, 1.8, 23.2)	G(1.4, 2.2, 4.7)	N(1.9, 0.9)	3.98	2.18

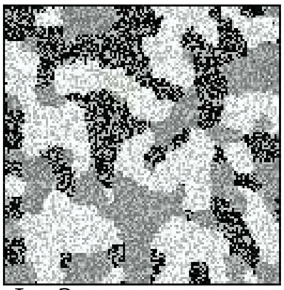
Figure 1

An example unsupervised segmentations of two sensor hidden evidential Markov field image.



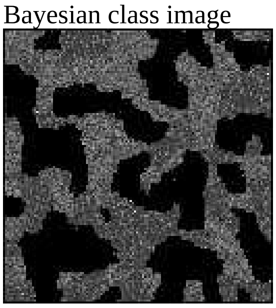
Im 1

Bayesian class image



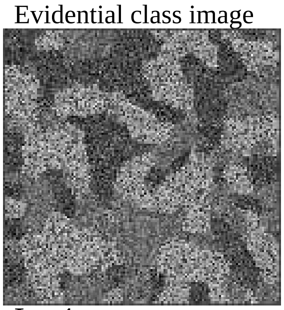
Im 2

Evidential class image



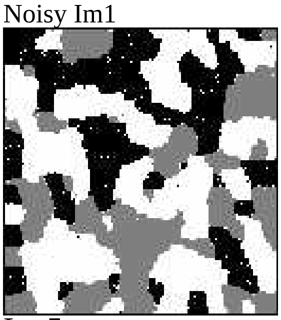
Im 3

Noisy Im1



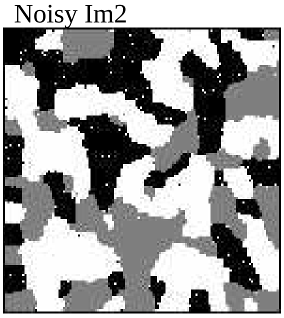
Im 4

Noisy Im2



Im 5

Unsupervised segmentation of Im3 with MPM, case 4, Table 1.



Im 6

Unsupervised segmentation of (Im3, Im4) with MPM after fusion, case 4, Table 1.

Concerning the detection of the distribution forms, we note that ICE-KERNEL finds the right forms and ICE-PEAR makes one mistake. It is interesting to note that ICE-based segmentation, which is necessarily based on normal distributions, gives here better results than the ICE-PEAR-based segmentation.

Finally, we note that the ICE-KERNEL-based fusioned MPM segmentation is close to the True Laws based fusioned MPM segmentation.

7. Conclusion

We have addressed in this paper the problem of parameter estimation in recent hidden evidential Markov fields model [4], with application to segmentation of multisensor images. Adopting the modelling by mass functions, we have considered the parameter estimation problem as a classical mixture estimation problem. We then proposed an original generalized mixture estimation method, which belongs to the wide family of methods proposed in [9], and we have applied it to solve the parameter estimation problem in the context of the multisensor evidential Markovian field model considered. First simulations show favourable behavior of the unsupervised MPM segmentation, i.e., the estimated parameter based estimation is close to the true parameter estimation.

We have considered a relatively simple case, with one sensor Bayesian and the another one Evidential. As a direction for further work, one could consider the problem of identifying of the mass functions related to each sensor, which will undoubtedly take a great importance in real situations.

References

[1] A. Appriou, Probabilités et incertitude en fusion de données multisenseurs, *Revue Scientifique et Technique de la Défense*, No. 11, pp. 27-40, 1991.

[2] J. Besag, On the statistical analysis of dirty pictures, *Journal of the Royal Statistical, Society*, Series B, 48, pp. 259-302, 1986.

[3] A. Bendjebbour and W. Pieczynski, Segmentation d'images multisenseur par fusion évidentielle dans un contexte markovien, *Traitement du Signal*, Vol. 14, No. 5, pp. 453-464, 1997.

[4] A. Bendjebbour and W. Pieczynski, Multisensor Evidential Hidden Markov Fields and Image Segmentation, 2nd IEEE International Conference on Intelligent Processing Systems (ICIPS'98), Gold Coast, Australia, August 4-7, 1998.

[5] R. Chellapa and A. Jain Ed., Markov Random Fields, *Academic Press*, 1993.,

[6] R.C. Dubes and A.K. Jain, Random field models in image analysis, *Journal of Applied Statistics*, Vol. 16, No 2, 1989.

[7] Y. Delignon, A. Marzouki, W. Pieczynski, Estimation of Generalized Mixture and Its Application in Image Segmentation, *IEEE Transactions on Image Processing*, Vol. 6, No. 10, pp. 1364-1375, 1997.

[8] S. Geman, and G. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on PAMI*, Vol. 6, pp. 721-741, 1984.

[9] N. Giordana and W. Pieczynski, Estimation of Generalized Multisensor Hidden Markov Chains and Unsupervised Image Segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 5, pp. 465-475, 1997.

[10] N. Giordana and W. Pieczynski, Chaînes de Markov cachées et restauration de données multicapteur, *Revue de Statistique Appliquée*, to appear.

[11] J. Guan and D.A. Bell, Evidence Theory and Its Applications, North-Holland, Amsterdam, 1991.

[12] G. J. Klir and T. A. Folger, Fuzzy sets, uncertainty, and information, Prentice Hall, 1988.

[13] S. Lakshmanan, and H. Derin, Simultaneous parameter estimation and segmentation of Gibbs random fields, *IEEE Transactions on PAMI*, Vol. 11, pp. 799-813, 1989.

[14] J. Marroquin, S. Mitter, and T. Poggio, Probabilistic solution of ill-posed problems in computational vision, *Journal of the American Statistical Association*, 82, pp. 76-89, 1987.

[15] W. Pieczynski, Champs de Markov cachés et estimation conditionnelle itérative, *Traitement du Signal*, Vol. 11, No. 2, pp. 141-153, 1994.

[16] G. Shafer, A mathematical theory of evidence, Prinston University Press, Prinston, 1976.

[17] P. Smets, Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem, Int. Jour. Approx. Reason., 9, pp. 1-35, 1993.

[18] R. R. Yager, J. Kacprzyk, and M. Fedrizzi, Advances in the Dempster-Shafer theory of evidence, Wiley, 1994.

[19] L. Younes, Estimation and annealing for Gibbsian fields, *Annales de l'Institut Henri Poincaré*, Vol. 24, No. 2, 1988, pp. 269-294.

[20] L. Younes, Parametric inference for imperfectly observed Gibbsian fields, *Probability Theory and Related Fields* 82, 1989, pp. 625-645.