STATISTICAL IMAGE SEGMENTATION

Wojciech Pieczynski
Groupe Image
Département Systèmes et Réseaux
Institut National des Télécommunications
9, rue Charles Fourier 91011 Evry Cedex France

Abstract. This work deals with the unsupervised Bayesian segmentation of images. We describe a recent general procedure of estimation of components in a mixture of distributions and show some possibilities of its application to the problem of the unsupervised segmentation. As far as modelling by Markov random fields is concerned the procedure presented can generate several particular algorithms of unsupervised segmentation of images.

1. Introduction

Our work deals with the problem of the unsupervised segmentation of images whose importance has been pointed out by several authors ([8], [9], [21]). When considering the statistical approach, with Bayesian segmentation, the modelling by Markov random fields is the generally adopted one. When distributions of corresponding Markov fields are sufficiently known there exist some efficient algorithms (simulated annealing, MPM, ICM) of Bayesian segmentation. When they are not sufficiently known the problem becomes more difficult: one has to estimate the useful parameters, either in a previous step or simultaneously. The statistical problem is that of "lost data": parameters are estimated from random variables whose distribution is a mixture of distributions.

The aim of this paper is to show how a recent procedure of estimation in the case of lost data, whose generality goes beyond the problems of image processing, can generate numerous algorithms of unsupervised Bayesian segmentation. The development of this paper is as follows:

In the following section we recall the modelling by Markov random fields and three Bayesian methods of segmentation.

The third section is devoted to a recent iterative method of estimation, called iterative conditional estimation (ICE). We introduce it in the case of a finite mixtures, sufficient to deal with the problem of segmentation, but a more general framework can be considered.

In the fourth section we show how the ICE procedure can generate numerous algorithms of unsupervised Bayesian segmentation and give an example.

Machine GRAPHICS and VISION vol. 1, nos. 1/2, 1992, pp. 261–268
(Proceedings of 2nd GKPO'92 Conference, Nałęczów, May 18–23, 1992)

The final section comprises the conclusion.

2. Modelling and Bayesian Segmentation

When considering the statistical segmentation of images authors generally suppose the existence of two random fields: the field of "classes" $\zeta = (\zeta_s)_{s \in S}$, and the field of "measurements" $X = (X_s)_{s \in S}$. Each ζ_s takes its values in a finite set $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ of classes and X_s in R^d, with $d \in N$. So the problem of segmentation in the problem of the estimation of an "ignored" realization of ζ from an "observed" realization of X. We will suppose that realizations of X depend on realizations of ζ and a noise containing as well the "natural variability" or "texture" as the disturbances due to the "transmission". The distribution of (ζ, X) is defined by P_ζ, distribution of ζ, and the family P_X^ε of distributions of X conditional to $\zeta = \varepsilon$. The field ζ will be supposed Markovian with all realizations possible; so its distribution is Gibbs distribution:

$$P[\zeta = \varepsilon] = P_\zeta[\varepsilon] = ke^{-U_\alpha(\varepsilon)} \tag{1}$$

where k is unknown, α a parameter and the form of U_α simple enough to make possible the calculation of the spatially conditional distributions. The distribution of X can be defined by m distributions $P^1, P^2, \ldots, P^m$ of X conditional to m uniform realizations ($\zeta_s = \omega_1$ for each s, $\zeta_s = \omega_2$ for each $s, \ldots, \zeta_s = \omega_m$ for each s respectively) and the hypothesis according to which the random variables X_i (with X_i restriction of X to $Q_i = \{s \in S / \zeta_s = \omega_i\}$) are independent. In order to allow simulations of $P^1, P^2, \ldots, P^m$ we will suppose that they are Gibbs distributions.

In the most general case Bayesian segmentation consists in estimating the realization of ζ_A from X_B, for $A, B \subset S$. More precisly ζ_A is estimated by ζ_A which maximises the "posterior" (conditional to $X_B = x_B$) distribution of ζ_A. Different choices of A, B generate different approaches: our work concerns the "global methods" which correspond to $B = S$ (pixels are classified using the whole information simultaneously), "local methods" corresponding to B of "small size" with $A \subset B$ (each pixel is classified from information contained in its neighbourhood). The best known global approaches are:

- MAP corresponds to $A = S$, $B = S$. The problem is resolved by Geman and Geman's ([11]) iterative algorithm (simulated annealing).
- MPM corresponds to $A = \{s\}$, $B = S$. In this case the problem can be treated by Marroquin et all's algorithm ([16]).
- ICM. This derministic algorithm of J. Besag ([2]) is rather a local one but we will range it with global methods because it necessites the same modelling (using Gibbs distributions) that MAP or MPM.

Thus, we can apply one of the above-mentioned methods if we know the corresponding parameters. Let us denote α parameters concerning the distribution of ζ and β parameters corresponding to the conditional distributions of X. As we said above, α is a parameter, generally appearing in the function "energy" U of the Gibbs distribution (1), knowning which makes it possible to compute the distribution of each ζ_s conditional to $(\zeta_t)_{t \neq s}$. The use of global methods does not allow the taking of the noise correlation into account: the random variables (X_s) are assumed to be independent conditionally to each realization of ζ. So, if we have m classes, it is sufficient to define m distributions of each X_s (conditional to $\zeta_s = \omega_1, \ldots, \omega_m$ respectively). Thus β will have m components $\beta_1, \ldots, \beta_m$. For instance, if the field X is Gaussian conditionally to ζ and each X_s real, we have $\beta_i = (\mu_i, \sigma_i^2)$, where μ_i is the mean and σ_i^2 the variance of the Gaussian distribution of each X_s conditional to $\zeta_s = \omega_i$.
In practice we just often have an observation $X = x$.

Thus we have to choose a segmentation method and propose an estimator for the corresponding parameter (α, β). The distribution of X is:

$$P_x = \sum_{\varepsilon} P_\zeta[\varepsilon] P^\varepsilon \tag{2}$$

which is said to be a "finite mixture of distributions". The problem of estimating (α, β) (in a certain sense "components of the mixture") is a very general statistical problem which often arises when dealing with Bayesian classification. We describe in the next section a recent general procedure of estimation and describe, in section 4, how this procedure can generate several agorithms of unsupervised segmentation.

3. Iterative Conditional Estimation (ICE)

Let us consider, in a general manner, a couple of random variables (ζ, X), ζ taking its value in a finite set and X in R^d. Let us suppose that the distribution of (ζ, X) is defined by P_ζ, distribution of ζ, and the family P_X^ε (ε being a possible realization of ζ) of distributions of X conditional to $\zeta = \varepsilon$. The distribution P_ζ depends on a parameter α and the family P_X^ε on a parameter β (the correspondence between the distributions of (ζ, X) and (α, β) is not necessarily bijective). X is the only observable and the problem is to estimate the parameter (α, β).

Let us suppose that we have at our disposal two estimators $\hat{\alpha} = \hat{\alpha}(\zeta)$ and $\hat{\beta} = \hat{\beta}(\zeta, X)$, $\hat{\alpha}$ being defined from ζ and $\hat{\beta}$ from (ζ, X). The direct use of $\hat{\alpha}$ and $\hat{\beta}$ is impossible because realizations of ζ are ignored, so, we have to approach $\hat{\alpha}$ and $\hat{\beta}$ by some function of X, the only one observable. The best approximation, as far as the square error is concerned, is the conditional expectation. To be more precise, if, for two random variables Y, Z, we denote the conditional expectation by $E[Y/Z]$ we can state:

$$E[(Y - E[Y/Z)^2] = \min_{\varphi} E[(Y - \varphi(Z))^2] \tag{3}$$

Thus we are led to condition $\hat{\alpha}$ and $\hat{\beta}$ by X: in doing this we lose the quality as "estimators". In fact, $\hat{\alpha}$ is independent from the parameter (α, β) but $E[\hat{\alpha}/X]$, which, for each $X = x$ is the expectation of $\hat{\alpha}$ according to the distribution of ζ conditioned by $X = x$ (so called "posterior" distribution) depends on (α, β). The same is true for $E[\hat{\beta}/X]$. Hence, if we want to compute $E[\hat{\alpha}/X = x]$ and $E[\hat{\beta}/X = x]$ we have to take a parameter (α_n, β_n) previously defined by some way. This defines an iterative procedure wich will be called Iterative Conditional Estimation (ICE, [20]). By denoting E_n the conditional expectation using $(\alpha_n, \beta_n$ the procedure is:

- one takes an initial value (α_0, β_0)
- $(\alpha_{n+1}, \beta_{n+1})$ is computed from (α_n, β_n) and X=x by:

$$\alpha_{n+1} = E_n[\hat{\alpha}/X = x] \tag{4}$$
$$\beta_{n+1} = E_n[\hat{\beta}/X = x] \tag{5}$$

In order to simplify matters let us denote by P_n^x the distribution of ζ conditioned by $X = x$ (posterior distribution) computed from (α_n, β_n). (4) and (5) can be written:

$$\alpha_{n+1} = \sum_{\varepsilon} \hat{\alpha}(\varepsilon) P_n^x[\varepsilon] \tag{6}$$
$$\beta_{n+1} = \sum_{\varepsilon} \hat{\beta}(\varepsilon, x) P_n^x[\varepsilon] \tag{7}$$

(4) and (5) or (6) and (7) are not necessarily workables in practice. As we will see in next section, P_n^x is not entirely known when the problem of image segmentation is tacled from modelling by Markov random fields, but it is possible to simulate realizations of ζ according to P_n^x. In situations like this we can use an "approximative" ICE: (6) and (7) can be approached, in accordance with the law of large numbers, by:

$$\alpha_{n+1} = \frac{1}{N} \sum_{i=1}^{N} \hat{\alpha}(\varepsilon_i) \tag{8}$$
$$\beta_{n+1} = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}(\varepsilon_i, x) \tag{9}$$

where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ are realizations of ζ according to P_n^x.
Finally, we can use the ICE procedure as soon as we get:

- An estimator $\hat{\alpha} = \hat{\alpha}(\zeta)$ of α and another one $\hat{\beta} = \hat{\beta}(\zeta, X)$ of β.
- An initial value (α_0, β_0) not "too far" from the real parameter (α, β).
- A way of computing $E_n[\hat{\alpha}/X]$ and $E_n[\hat{\beta}/X]$ or a way of simulating realizations of ζ according to the posterior distribution P_n^x.

4. Unsupervised Segmentation of Images

The ICE procedure can generate a great number of non supervised segmentation algorithms (NSSA). α is a parameter, generally appearing in the energy of the Gibbs distribution of ζ, knowing which allows the computing of the conditional distributions. β is a parameter defining the distributions, conditional to $\zeta = \varepsilon$, of X. The use of ICE is possible as soon as we can simulate realizations of ζ according to the posterior distribution, the latter being computed with the current parameter (α_k, β_k). Thus, the possibility of simulations requires the markovianity of the posterior distribution which is ensured by the conditional independence of the random variables (X_s). Let us suppose X real and conditionally Gaussian. So the parameter β is $\beta = (\beta_1, \ldots, \beta_m)$, wich m the number of classes, and $\beta_i = (\mu_i, \sigma_i^2), \mu_i, \sigma_i^2$ being respectively the mean and variance of the distribution of each X_s conditional to $\zeta_s = \omega_i$ $(for\ 1 \leq i \leq m)$.

So ICE will define a NSSA as soon as we take $\hat{\alpha}, \hat{\beta}$ estimators of α, β respectively (from (ζ, X)) and choose between MAP, MPM, ICM. Let us mention four possibilities for $\hat{\alpha}$:

$\hat{\alpha}_1$: pseudo-maximum likelihood ([2]). The idea is to substitute the likelihood(1), where k is unknown, by the product of conditional distribution. α is then estimated by $\alpha^\star$ which, for a given realization of ζ, maximises the function thus obtained.

$\hat{\alpha}_2$: least square error method. The description of this method, proposed by Derin and Elliot, can be seen in [9].

$\hat{\alpha}_3$: maximum likelihood. L. Younes proposes ([26]) an iterative procedure defining a sequence converging to the maximum of the likelihood defined by (1). His algorithm, based on the stochastic gradient, allows a rigorous mathematical justification.

$\hat{\alpha}_4$: algorithm of Metropolis. Its description can be seen in[13].

The more natural choice of the estimator $\hat{\beta} = \hat{\beta}(\zeta, X)$, in the hypothesis when conditional distributions of X are Gaussian, are the empirical mean and variance (μ_i, σ_i^2) which are defined from $s \in S$ such that $\zeta_s = \omega_i$. This can be written:

$$\hat{\mu}_i = \frac{\sum_{s \in S} X_s 1_{[\zeta_s = \omega_i]}}{\sum_{s \in S} 1_{[\zeta_s = \omega_i]}} \tag{10}$$

$$\hat{\sigma}_i^2 = \frac{\sum_{s \in S} (X_s - \hat{\mu}_i)^2 1_{[\zeta_s = \omega_i]}}{\sum_{s \in S} 1_{[\zeta_s = \omega_i]}} \tag{11}$$

Formulae (4) and (5) are not workable and have to use (6) and (7) the sequence $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ being simulated by the Gibbs sampler ([10]). Then we can regulate the importance of the :"stochastic" aspect of ICE by choosing N: when N increases the stochastic side of the algorithm decreases. The intentional choice of N "small" can increase its efficiency: the authors of the SEM algorithm([4]), wich can be seen as a particular case of ICE with N=1 and the choice of (8) instead of (4), show its superiority over the EM algorithm in a simple case of a mixture of Gaussian densities. It is also possible to make N depend on the number k of iteration and introduce a kind of "cooling".

When combining these different possibilities we arrive twelve global unsupervised methods of segmentation. For instance, taking α_3, β and MAP the proceeding is:

- one takes $\Theta_0 = (\alpha_0, \beta_0)$
- $\Theta_{k+1} = (\alpha_{k+1}, \beta_{k+1})$ is computed from $\theta_k = (\alpha_k, \beta_k)$ and X=x in the following way:
  - using the Gibbs sampler N realizations $\varepsilon_1, \varepsilon_2, ..., \varepsilon_N$ of ζ are simulated according to the posterior distribution computed from (α_k, β_k) and x,
  - for each ε_j the parameter α is estimated by the iterative algorithm of L. Younes (wich requires m new simulations by the Gibbs sampler) – let us denote $\alpha(\varepsilon_j)$ the result obtained,
  - for each ε_j the parameter β is estimated by formulae (10) and (11), which gives $\beta(\varepsilon_j)$,
  - $\alpha_{k+1}, \beta_{k+1}$ are obtained from $(\alpha(\varepsilon_j), \beta(\varepsilon_j))$, $\quad 1 \leq j \leq N$ by (8) and (9),
- if the sequence (α_k, β_k) becomes steady, the estimation step is finished and one proceeds to the segmentation using the simulated annealing([10]).

A comparative study between this algorithm and the simulated annealing based on parameters estimated by an another algorithm of L. Younes ([27]) can be seen in [1].

5. Conclusion

We presented some applications of the ICE procedure to unsupervised image segmentation. We have limited our investigations to global methods but the same ICE general method can also generate some local ones ([3],[14],[15],[20]). It is therefore possible to define numerous NSSA: probably none of them is better than any other in the general case. From theoretical point of view the study of their asymptotical behaviour us without doubt, rather tedious: until now we have only had some partial results([5],[13],[19],[26]). Numerous publications concerning the EM algorithm, in the case of independent data, lead us to imagine the extent of the difficulty.

From the point of view of applications there is a wide study of tests to do: possibilities of applications are extremly various and the problem is that of knowing what family of NSSA is well suited to given kind of images. We have begun some studies ([3],[14]) based on simulations of different kinds of noises but the collaboration of experts in the processing of real images would be, undoubtedly, of the utmost importance.

References

[1] O. Allagnat, J.M. Boucher, D.C. He, W. Pieczynski, Hidden Markov fields and unsupervised segmentation of images, IAPR 11th International Conference on Pattern Recognition (ICPR 92), Delft, Pays Bas, September 92.

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

[3] B. Braathen, N. Marhic, P. Masson, W. Pieczynski, Sur une nouvelle approche de segmentation Bayesienne non supervisée d'Images– Treisième colloque GRETSI, Juans-les-Pins, France, September 1991.

[4] H. Caillol, W. Pieczynski, Fuzzy statistical unsupervised Bayesian image segmentation, IEEE International Geoscience and Remote Sensing Symposium (IGARSS 92), Houston, USA, May 1992.

[5] G. Celeux, J. Diebolt, L'algorithme SEM: un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densites, Revue de Statistique Appliqee, Vol. 34, 2 (1986).

[6] B. Chalmond, An iterative Gibbsian technique for reconstruction of m-ary images, Pattern Recognition, Vol 22, 6 (1989), 747–761.

[7] M.M. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society, Series B, 39 (1977), 1–38.

[8] P.A. Devijver, M. Dekesel, Champs aléatoires de Pickard et modélisation d'images digitales, Traitement du Signal, Vol. 5, 5 (1988).

[9] R.C. Dubes, A.K. Jain, Random field models in image analysis, Journal of Applied Statistics, Vol. 6, 2 (1989).

[10] M. Emsalem, H. Caillol, P. Olivie, G. Carnat, W. Pieczynski, Fast unsupervised statistical image segmentation, IEEE International Geoscience and Remote Sensing Symposium (IGARSS 92), Houston, USA, May 1992.

Machine GRAPHICS and VISION vol. 1, nos. 1/2, 1992, pp. 261–268
(Proceedings of 2nd GKPO'92 Conference, Nałęczów, May 18–23, 1992)

[11] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Transactions on PAMI 6, 721–741, 1984.

[12] R. Haralick, J. Hyonam, A context classifier, IEEE Transactions on GRS-GE-24, 997–1007, 1986.

[13] P.A. Kelly, H. Derin, K.D. Hartt, Adaptive segmentation of speckled images using a hierarchical random field model, IEEE Transations on ASSP, Vol. 36, 10 (1988), 1628–1641.

[14] N. Marhic, P. Masson, W. Pieczynski, Melange de lois et segmentation non supervisee des donnees SPOT, to appear in Statique et Analyse des Donnees, 1992.

[15] N. Marhic, W. Pieczynski, Estimation of mixture and unsupervised segmentation of images, IEEE International Geoscience and Remote Sensing Symposium (IGARSS 91), Helsinki, Finland, May 1991.

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

[17] P .Masson, W. Pieczynski, Segmentation of SPOT images by contextual SEM, Proceedings of EUSIPCO, Barcelone, Spain, November 1990.

[18] E. Mohn, N. Hjort, G. Storvic, A simulation study of some contextual classification methods for remotely sensed data, IEEE Transactions on GRS-GE-25, 796–804, 1987.

[19] W. Pieczynski, Estimation of context in random fields, Journal of Applied Statistics, Vol. 16, 2 (1989).

[20] W. Pieczynski, Mixture of distributions, Markov random fields and unsupervised bayesian segmentation of images, Technical Report No 122, L.S.T.A., Universite Paris VI, 1990.

[21] W. Qian, D.M. Titterington, On the use of Gibbs Markov chain models in the analysis of images based on second-order pairwise interactive distributions, Journal of Applied Statistics, Vol. 16, 2 (1989).

[22] H.C. Quelle, J.M. Boucher, W. Pieczynski, Unsupervised Bayesian classification of SAR images, IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Helsinki, Finland, May 1991.

[23] P. Swain, S. Vardeman, J. Tilton, Contextual classification of multispectral image data, Pattern Recognition 13, 429–441, 1981.

[24] J. Tilton, S. Vardemann, P. Swain, Estimation of context for statistical classification of multispectral image data, IEEE Transactions on GRS-GE-20, 445–452, 1982.

[25] A. Veijanen, A simulation-based estimator for hidden Markov random fields, IEEE Transactions on PAMI, Vol. 13, 8 (1991).

[26] L. Younes, Estimation and annealing for Gibbsian fields, Annales de l'Institut Henri Poincaré 2, 1988.

[27] L. Younes, Parametric inference for imperfectly observed Gibbsian fields, Probability Theory and Related Fields 82, 625–645, 1989.

Machine GRAPHICS and VISION vol. 1, nos. 1/2, 1992, pp. 261–268
(Proceedings of 2nd GKPO'92 Conference, Nałęczów, May 18–23 , 1992)