ESTIMATION OF THE WEIGHT PARAMETER WITH SAEM FOR MARKED POINT PROCESSES APPLIED TO OBJECT DETECTION

Aurélie Boisbunon[1,2], *Josiane Zerubia*[1]

[1]INRIA, AYIN research group, 2004 Route des Lucioles, BP 93, 06902 Sophia-Antipolis Cedex, France
{aurelie.boisbunon, josiane.zerubia}@inria.fr
[2] CNES, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France

ABSTRACT

We consider the problem of estimating one of the parameters of a marked point process, namely the tradeoff parameter between the data and prior energy terms defining the probabilit density of the process. In previous work, the Stochastic Expectation-Maximization (SEM) algorithm was used. However, SEM is well known for having bad convergence properties, which might also slow down the estimation time. Therefore, in this work, we consider an alternative to SEM: the Stochastic Approximation EM algorithm, which makes an efficient use of all the data simulated. We compare both approaches on high resolution satellite images where the objective is to detect boats in a harbor.

***Index Terms*—** Image processing, object detection, marked point process, Stochastic EM, Stochastic Approximation EM, .

1. INTRODUCTION

Object detection is an important problem in several areas of image processing, such as monitoring of populations of animals or plants in ecology, monitoring of vehicles, boats, or cargos in highways/harbours, or detection of roads in remote sensing. When the number of objects is unknown and possibly important (several hundreds or thousands) and when the objects can be modelled by a simple parametric shape, it is interesting to use marked point processes [1, 2]. This stochastic geometry methodology combines object-based methods with probabilistic approaches. It consists in defining the probability density of the configuration of objects from two terms: the data term using the knowledge given by the intensity of an image and assessing the quality of the estimated configuration for the image; the prior term introducing any preliminary knowledge on the shape of the objects or on the interaction between them (see Section 2). These two terms are weighted by a tradeoff parameter, which does not have a physical meaning in most cases and is usually difficult to compute analytically or numerically. Indeed, we are faced here with two issues: first, both the configuration and the parameter are unknown, resulting in an incomplete data case; second, the probability density of the configuration involves a normalizing constant that depends on the parameter, making the optimization problem harder to solve.

The problem of estimating this weight parameter has been studied rather scarsely, and very few works actually consider a valid method for that purpose, while finding a correct value is crucial for a good extraction. Among such works, Chatelain *et al.* (2009) [3] and Ben Hadj *et al.* (2010) [4] dealt with the first issue by using the Stochastic Expectation-Maximization (SEM) algorithm developed by Celeux & Diebolt (1985) [5], which consists in alternatively simulating the configuration and approximating its expected likelihood for a fixed value of the parameter, and maximizing the likelihood for a fixed configuration (see Section 3.1). The second issue was treated by replacing the likelihood by the pseudo-likelihood, whose normalizing constant is much easier to compute. However, it is well known that the SEM algorithm only converges in law, and not pointwisely [6]. Therefore, it can lead to a prohibitive computational time for large images. This is the reason why, in this work, we investigate alternatives to SEM that have good convergence properties and can give an estimate of the parameters in a shorter time. In particular, we consider the Stochastic Approximation Expectation-Maximization (SAEM) algorithm proposed by Delyon *et al.* (1999) [7] (see Section 3.2). Section 4 shows numerical comparisons for the application on boat detection in harbours. Finally, we discuss in Section 5 of other possible methods that we intend to study in future works.

2. BASIC BACKGROUND ON MARKED POINT PROCESSES

Marked point processes add a mark on the point process representing the shape of the objects [1, 2, 8]. This shape should be characterized by a few number of parameters for an easier analysis and a faster computation. The probability density of the configuration of objects is modelled by a Gibbs distribu-

The first author would like to thank the French Spatial Agency (CNES) for the financial support of her post-doc at INRIA. Both authors thank Alain Giros from CNES for fruitful discussions.

tion

$$f_\theta[X|Y=\mathbf{y}] = \frac{1}{c(\theta)} \exp\{-U_\theta(X,\mathbf{y})\} \quad (1)$$

where X is the configuration of objects, $\mathbf{y}$ is the intensity of pixels in the image, $c(\theta)$ is the normalizing constant and $U_\theta(X,\mathbf{y})$ is the energy of the model. The latter term itself is decomposed as

$$U_\theta(X,\mathbf{y}) = \gamma_d U_\theta^d(X,\mathbf{y}) + U_\theta^p(X),$$

where $U_\theta^d(X,\mathbf{y})$ is the data term assessing how well the estimated configuration fits the image, $U_\theta^p(X)$ is the prior term including any preliminary information, and γ_d is parameter realizing the tradeoff between the two latter terms and is an element of θ. Examples of U_d are functions of the contrast between the intensity in the interior the object and the one in its border, measured by a distance. The prior term can take into account the specificities of the distribution of the objects in the space, such as orientation, alignment, overlapping ... Hence, the parameter θ includes not only the weight parameter γ_d, but also thresholds on the contrast and the measure of interactions.

In the numerical study of Section 4, we will focus on the ellipse model developed by Chatelain *et al.* (2009) [3] and extended both by Ben Hadj *et al.* (2010) [4] and Craciun & Zerubia (2013) [9] for boat detection in harbors. In this model, which was adapted from the one for tree detection by Perrin *et al.* (2004) [10], they used the Bhattacharya distance for the contrast, and Craciun & Zerubia (2013) [9] considered only alignment and relaxed orientation from the former model in [3, 4], while using a more sophisticated model of ellipses.

3. ESTIMATION OF THE WEIGHT PARAMETER

The estimation of γ_d can be described by the following optimization problem

$$\max_{\gamma \in \mathbb{R}_+^*} \{l_\gamma(X,\mathbf{y}) = f_\theta[X|Y=\mathbf{y}]\}, \quad (2)$$

which corresponds to the maximization of the likelihood, that is, we look for the value of γ_d yielding the most probable configuration. As mentioned in the introduction, this problem is hard to solve for two reasons: (i) the problem corresponds to an incomplete data case since both X and γ_d are unknown and (ii) the normalizing constant $c(\theta)$ depends on all the parameters, including the weight γ_d. Also, it belongs to the set $(0,\infty)$ so that estimation from a grid or from trial and error cannot be efficient. Up to now, the answer to both problems has been given by the Stochastic EM algorithm applied to the pseudo-likelihood, which we develop in the next paragraph.

Note that the other parameters (thresholds on contrast, orientation, or weight between the different terms involved in the prior) have been chosen by trial and error sofar, but will also require full consideration in future works.

3.1. Pseudo-likelihood and Stochastic EM

The Stochastic Expectation-Maximization (SEM) algorithm was designed by Celeux & Diebolt (1985) [5] as an alternative to EM for cases where the exact expected likelihood can be difficult to compute, either analytically or numerically. The expectation step is thus replaced by a simulation step followed by an approximation of the expectation based on the simulated data.

The algorithm cannot be applied directly here, since the normalizing constant in the likelihood is itself hard to compute. Therefore, Chatelain *et al.* (2009) [3] proposed to replace it by the pseudo-likelihood as was done by Besag (1975) [11]

$$\begin{aligned} PL_\theta(\mathbf{x},\mathbf{y}) &= \exp\left\{-\int_{\mathcal{X}\times\mathcal{K}} \lambda_\theta(u;\mathbf{x},\mathbf{y})\Lambda(du)\right\} \\ &\quad \times \left\{\prod_{x_i\in\mathbf{x}} \lambda_\theta(x_i;\mathbf{x},\mathbf{y})\right\} \end{aligned} \quad (3)$$

where $\lambda_\theta(u;\mathbf{x},\mathbf{y}) = \beta\exp\{-\gamma_d U_d(u) - U_p(u)\}$ is the extended Papangelou intensity for one object and Λ is the Poisson distribution. In other words, the pseudo-likelihood considers all objects to be independent. Note that the first term in Equation (3) represents the normalizaing constant of the pseudo-likelihood and its computation can be done numerically much more easily than the likelihood in Equation (1). More details can be found in [4].

The SEM algorithm hence gives an approximation of the maximum of pseudo-likelihood (MPL) estimator, as described in Algorithm 1. However, the SEM algorithm presents a major issue: it does not converge pointwisely, as mentioned by Celeux *et al.* (1996) [6].

Algorithm 1 SEM algorithm for Problem (2)

Inputs

- Initial value $\gamma^{(0)}$, image $\mathbf{y}$

$k = 0$

repeat

(S) Simulate $X^{(k)} \sim l_{\gamma^{(k)}}(X,\mathbf{y})$

(E) Compute an approximation of the expected pseudo-likelihood

$$\hat{Q}(\gamma,\gamma^{(k)};\mathbf{y}) = \log PL_\theta(\mathbf{x}^{(k)},\mathbf{y})$$

(M) Update γ with

$$\gamma^{(k+1)} = \arg\max_\gamma \hat{Q}(\gamma,\gamma^{(k)};\mathbf{y})$$

$k \to k+1$

until $|\gamma^{(k+1)} - \gamma^{(k)}| \leq \epsilon$

3.2. Improving on the convergence of SEM

3.2.1. Monte Carlo EM and Simulated Annealing

In order to overcome the issues of convergence of SEM, Celeux *et al.* (1996) [6] suggested averaging over the next iterations. This amounts to running the Monte Carlo Expectation-Maximization (MCEM) algorithm for the last step. Indeed, the SEM algorithm is actually a particular case of the MCEM algorithm, developped by Wei & Tanner (1990) [12], where several observations $\mathbf{x}_1, \ldots, \mathbf{x}_M$ are generated at the simulation step and the expectation step is obtained by averaging the log-likelihood over the M observations. Hence, SEM is just MCEM with $M = 1$. However, such a strategy raises the question of how many observations should we generate in order to obtain a satisfying value of γ_d. Especially, in our context where $\mathbf{x}$ is the configuration of all the objects, each observation is obtained by running of a Multiple Birth and Death (MBD) algorithm on the current configuration [2]. This means we have to run enough moves of the MBD algorithm for the new configuration to be different from the current one. Hence, a single simulated observation already results from a large number of simulated moves from MBD, and the computational cost of m observations can be prohibitive even for a moderate number M.

On the other hand, Ben Hadj *et al.* (2010) [13] and Descombes (2013) [2, Chapter 7] considered running a Simulated Annealing algorithm after few iterations of SEM in order to reach convergence. In other words, the maximization step is modified by

$$\gamma^{(k+1)} = \begin{cases} \arg\max_\gamma \hat{Q}(\gamma, \gamma^{(k)}; \mathbf{y}) & \text{if } u \leq e^{\frac{\hat{Q}^{(k+1)} - \hat{Q}^{(k+1)}}{T_k}} \\ \gamma^{(k)} & \text{otherwise,} \end{cases}$$

where u is generated from the uniform distribution on $[0; 1]$, $\hat{Q}^{(k+1)} = \arg\max_\gamma \hat{Q}(\gamma, \gamma^{(k)}; \mathbf{y})$ and $T_k = T_0 / \log(k+1)$ is the temperature of the Simulated Annealing process. In the sequel, we refer to that procedure as SEM-SA. Although this ensures a better estimate of γ_d, it also increases the computation time compared to SEM.

3.2.2. Stochastic Approximation EM

In view of the computational limitations of the marked point process we consider, we propose to deal with another option, namely the Stochastic Approximation Expectation-Maximization (SAEM) algorithm. The SAEM algorithm was proposed by Delyon *et al.* (1999) [7] as alternative to the Monte Carlo and the Stochastic EM algorithms. SAEM itself differs from both algorithms in the way the expectation is approximated. Indeed, it is based on the Stochastic Approximation method introduced by Robbins & Monro (1951) [14], which in our context corresponds to

$$\begin{aligned} \hat{Q}(\gamma, \gamma^{(k)}; \mathbf{y}) &= (1 + \tau_k)\hat{Q}(\gamma, \gamma^{(k-1)}; \mathbf{y}) \\ &\quad + \frac{\tau_k}{M(k)} \sum_{i=1}^{M(k)} \log PL_\theta(\mathbf{x}_i^{(k)}, \mathbf{y}), \end{aligned}$$

where τ_k is the step size, also called forgetting factor. Hence, it keeps memory of the past simulations through a convex combination of the previous approximated expectation and the one for the new simulated data governed by the stepsize. Jank (2006) [15] argues that this stepsize is crucial for the algorithm to have good convergence properties. In particular, he considers the choice $\tau_k = k^{-\alpha}$, with $\alpha \in (0.5, 1)$.

In our case, we consider the case where $M(k) = 1$, just as in SEM, for the reason we mentioned before about computational cost in the simulation scheme. Therefore, SAEM improves on both MCEM and SEM by taking the best of both worlds: good convergence properties of the approximation obtained from only one simulated observation. It should thus be faster than the SEM-SA procedure proposed in [2, Chapter 7] at little or no cost in accuracy of extraction.

Algorithm 2 SAEM algorithm for Problem (2)

Inputs

- Initial value $\gamma^{(0)}$, image $\mathbf{y}$

$k = 0$

repeat

(S) Simulate $X_m^{(k)} \sim l_{\gamma^{(k)}}(X, \mathbf{y})$, $m = 1, \ldots, M(k)$

(E) Compute an approximation of the expected pseudo-likelihood

$$\begin{aligned} \hat{Q}(\gamma, \gamma^{(k)}; \mathbf{y}) &= (1 + \tau_k)\hat{Q}(\gamma, \gamma^{(k-1)}; \mathbf{y}) \\ &\quad + \tau_k \log PL_\theta(\mathbf{x}^{(k)}, \mathbf{y}) \end{aligned}$$

(M) Update γ with

$$\gamma^{(k+1)} = \arg\max_\gamma \hat{Q}(\gamma, \gamma^{(k)}; \mathbf{y})$$

$k \to k+1$

until $|\gamma^{(k+1)} - \gamma^{(k)}| \leq \epsilon$

---

4. NUMERICAL RESULTS

We applied the SAEM algorithm to the program developed from the works in [3], [4] and [9] for the detection of boats on high resolution satellite images. The program mainly consists of two parts: first, the estimation of the parameters, which also give a first configuration of the objets as a result of the iterative algorithm ; then, a simulated annealing procedure for the refinement of the configuration based on the estimated parameters.

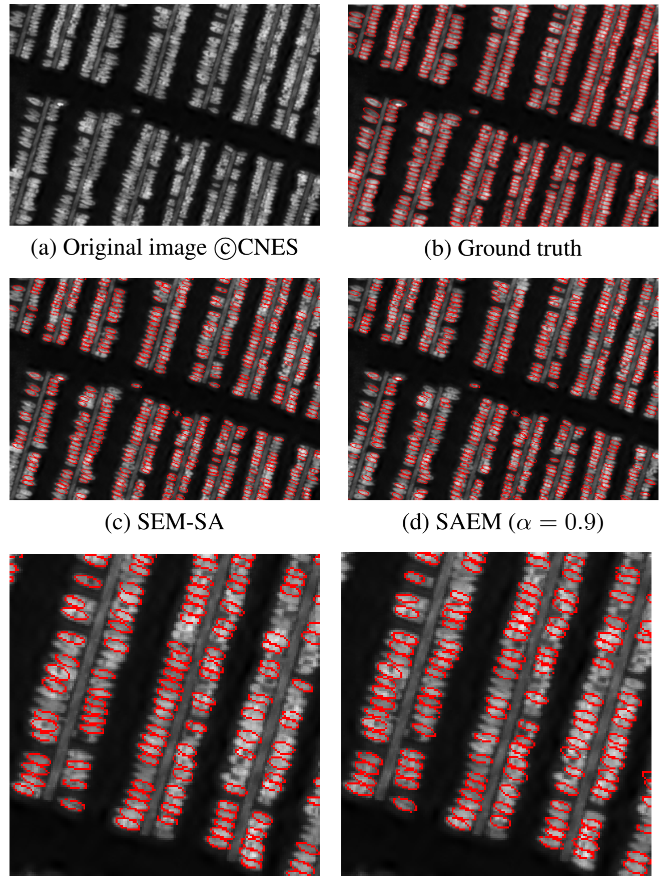
Before running the algorithm and comparing the results to those obtained with SEM, we did our own ground truth from a part of the original image, shown on Figures 1 (top).

We then run the program 50 times for different values of the forgetting factor $\alpha \in [0.5, 1)$, and computed several measures of performance. The first measure is the F-score which compares the quality of the extraction obtained at the end of the program compared to the ground truth. It is defined by

$$\text{F-score} = \frac{2TP}{2TP + FP + FN},$$

where TP (true positive) is the number of pixels correctly estimated as part of a boat, FP is the number of pixels that have been incorrectly estimated as part of a boat, and FN is the number of pixels that have been incorrectly estimated as part of the background. The F-score thus gives an estimation of the detection rate: the closer to 1, the better the detection is. The second measure is the time of computation of the estimation part: the program was run on Linux with a 2.4 GHz Intel processor and 4 GB of RAM. Note that the code is currently neither optimized nor parallelized and is mostly done in static programming.

Table 1 displays the F-scores and computational time of estimation averaged over the 50 runs, along with their standard deviations, for SAEM with different values of α compared with the SEM-SA procedure described in Section 3.2.1. We can first notice that SAEM does not lead to a loss in accuracy of the extraction since the difference in average detection rate with SEM-SA is less than 1% in the worst case ($\alpha = 0.8$), and about 0.3% in the best case ($\alpha = 0.9$). When looking at the time for computing the estimation part, we clearly see a gain of approximately 13% for $\alpha = 0.5$ or 0.9, whose average time is a little more than 7 minutes compared to the 8 minutes necessary for SEM-SA to give an estimate of γ_d. Hence, the advantage of using SAEM with $\alpha = 0.9$ is obvious here, as a similar detection rate is obtained in a shorter time. Note however that the gain might not be as important as expected. This is certainly due to the fact that the SEM-SA procedure is not programmed to reach convergence, but instead it is set to run the simulated annealing maximization step during only 10 iterations. The gain would probably be larger if this constraint was relaxed.



(a) Original image ©CNES (b) Ground truth

(c) SEM-SA (d) SAEM ($\alpha = 0.9$)

(e) SEM-SA, detail (f) SAEM ($\alpha = 0.9$), detail

Fig. 1. Detection of boats on a satellite image.

5. DISCUSSION

This work is a first step in the study of the estimation of the weight parameter. We proposed here to use the Stochastic Approximation EM algorithm instead of the procedure currently used for boat detection, which consists in an hybrid Stochastic EM algorithm where the maximization step is performed by Simulated Annealing. In practice, we were able to decrease the computational time for the estimation of the weight parameter at no loss in accuracy of the detection, which was our purpose. However, we believe that we can decrease even more the computational time with other methods. Therefore, in future work, we intend to study and compare methods such as Bayesian methods, Quasi Monte Carlo and Genetic Algorithms.

Finally, in the long run, it is important to check the validity and performances of these methods when we estimate more than one parameter, such as the weights for the different terms involved in the prior energy, or the thresholds on the contrast and the overlapping. This direction of research is however much more difficult to perform, as it is not clear how changing one parameter affects the others and as it can considerably increase the computational time.

References

[1] M. N. M. van Lieshout, *Markov Point Processes And Their Applications*, Imperial College Press, London, 2000.

[2] X. Descombes et al., *Stochastic geometry for image analysis*, Wiley/ISTE, 2013.

[3] F. Chatelain, X. Descombes, and J. Zerubia, "Param-

Forget factor α	**Detection rate** *(%)*		**Estimation time** *(min.)*	
0.5	70.04	(0.931)	**7.22**	(3.25)
0.6	69.94	(1.245)	7.55	(2.73)
0.7	69.96	(0.757)	8.23	(3.15)
0.8	69.91	(2.512)	7.73	(3.64)
0.9	**70.14**	(0.989)	7.31	(2.52)
SEM-SA	70.42	(1.199)	8.19	(2.01)

Table 1. Detection rate (%) and estimation time (min.) for different values of the forgetting factor in SAEM. The second and fourth columns correspond to mean values, and the third and fifth ones to standard deviations, all of them over 50 runs.

eter estimation for marked point processes. application to object extraction from remote sensing images.," in *Proc. Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, Bonn, Germany, August 2009.

[4] S. Ben Hadj, F. Chatelain, X. Descombes, and J. Zerubia, "Parameter estimation for a marked point process within a framework of multidimensional shape extraction from remote sensing images," in *Proc. ISPRS Technical Commission III Symposium on Photogrammetry Computer Vision and Image Analysis (PCV)*, Paris, France, September 2010.

[5] G. Celeux and J. Diebolt, "The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem," *Computational statistics quarterly*, vol. 2, no. 1, pp. 73–82, 1985.

[6] G. Celeux, D. Chauveau, and J. Diebolt, "Stochastic versions of the EM algorithm: an experimental study in the mixture case," *Journal of Statistical Computation and Simulation*, vol. 55, no. 4, pp. 287–314, 1996.

[7] B. Delyon, M. Lavielle, and E. Moulines, "Convergence of a stochastic approximation version of the EM algorithm," *The Annals of Statistics*, vol. 27, no. 1, pp. 94–128, 1999.

[8] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., 1987.

[9] P. Craciun and J. Zerubia, "Unsupervised marked point process model for boat extraction in harbors from high resolution optical remotely sensed images," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Melbourne, Australia, 2013.

[10] G. Perrin, X. Descombes, and J. Zerubia, "Tree crown extraction using marked point processes," in *Proc. European Signal Processing Conference (EUSIPCO), Vienna, Austria*, 2004.

[11] J. Besag, "Statistical analysis of non-lattice data," *The Statistician*, vol. 24, no. 3, pp. 179–195, 1975.

[12] G.C.G. Wei and M.A. Tanner, "A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms," *Journal of the American Statistical Association*, vol. 85, no. 411, pp. 699–704, 1990.

[13] S. Ben Hadj, F. Chatelain, X. Descombes, and J. Zerubia, "Estimation des paramètres de modèles de processus ponctuels marqués pour l'extraction d'objets en imagerie spatiale et aérienne haute résolution," Rapport de recherche 7350, INRIA, July 2010.

[14] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951.

[15] W. Jank, "Implementing and diagnosing the stochastic approximation EM algorithm," *Journal of Computational and Graphical Statistics*, vol. 15, no. 4, pp. 803–829, 2006.