UNSUPERVISED DETECTION OF CONTOURS USING A STATISTICAL MODEL

F. Destrempes ‡ *M. Mignotte* ‡

‡ DIRO, Département d'Informatique et de Recherche Opérationnelle, C.P. 6128,
Succ. Centre-Ville, Montréal, Canada (Quebec), H3C 3J7.
E-MAIL : DESTREMP@IRO.UMONTREAL.CA

ABSTRACT

In this paper, we describe an unsupervised segmentation method for contours which proves quite adapted for the images obtained by electronic acquisition. We present two statistical models for the norm of the gradient of the gray level at the pixels of an image, one for contour points and one for points outside contours. We also describe a Markov model with constraint which incorporates those two statistical distributions as likelihood together with a simple *a priori* model. Our model is suitable for an Iterative Conditional Estimation (ICE) procedure for the estimation of the parameters and an Iterated Conditional Modes (ICM) algorithm, or a simulated annealing, for the segmentation. A preliminary step proceeds to the segmentation of the image into sub-regions and uses a Markov model without constraint based on the gray level distribution on the image.

1. INTRODUCTION

Detection of contours is a fundamental problem in Image Processing. One popular tool for solving this problem is the Canny edge-detector or its variants (Deriche, Shen edge-detectors) based on the gradient of the gray level at the pixels of the image. The main inconvenient with these algorithms is the supervision in the specification of thresholds. Moreover, these algorithms are not based on the statistics of the image. In [1], a Markov model with constraints for segmentation of an image into regions or contours has been presented. However, again, the model is not directly based on the statistics of the image. In [2], an advance has been made in proposing a statistical model for the norm of the gradient of the gray level outside contours and for a related random variable on contours. However, we depart from the model presented in [2], since this model is adapted to a different problem.

Combining the ideas in [1] and [2], we present in this paper a Markov model that takes into account regions and contours as well as the statistics of the image. However, unlike [1], we split the procedure into two separate steps. The first step consists in a pre-segmentation of the image into classes of sub-regions. The gray level distribution on each class is modeled by a single Gaussian law. The estimation of the parameters is done by an ICE procedure and the ICM algorithm is used to obtain the segmentation. Taking contours of the segmented image, we obtain a subset T of the image that contains the contours of interest but also false contours which appear as artifacts. The second step consists in segmenting the points of T into two classes : the points off contours and the points on contours, thus removing the false contours. For this step, we model the norm distribution of the gradient of the gray levels by a Weibull law on the class "off" and by a mixture of three Gaussian laws on the class "on". We present a Markov model based on all pixels of the image subject to a constraint ; namely, the points outside T are imposed to be in the class "off". We use again an ICE procedure for the estimation of the parameters and the ICM algorithm for the segmentation. Despite its computational load, our method presents two advantages. Firstly, our method is totally unsupervised, contrary to simpler methods. Secondly, and more deeply, our model can be used to solve other problems, such as unsupervised localization of shapes, as will be shown elsewhere.

This paper is organized as follows. In Section 2, we present in details the pre-segmentation of the image into sub-regions. Section 3 presents the segmentation of the image into contours. In Section 4, we discuss briefly empirical results.

2. PRE-SEGMENTATION OF THE IMAGE INTO SUB-REGIONS

Given an image of size N, $G = (S, U)$ will denote the non-oriented graph consisting of the N pixels of the image with neighborhoods given by the usual 8-neighbors. We consider a couple of random fields $Z = (X, Y)$, where $Y = \{Y_s, s \in S\}$ represents the field of observations located on the sites s (associated to the pixels of the image), and $X = \{X_s, s \in S\}$ the label field.

In this section, Y_s represents the gray level at a pixel and takes its value in $\{0, \ldots, 255\}$, whereas X_s represents a class of sub-regions in the image and takes its values in $\{e_1, e_2, ..., e_K\}$, where $K = 5$ in all our experiments. The distribution of (X, Y) is defined by a prior distribution $P_X(x)$, supposed to be stationary and Markov and by site-wise likelihoods $P_{Y_s/X_s}(y_s/x_s)$ whose parameter Φ depends on the class label x_s. We assume independence between each random variable Y_s given X_s. The observable Y is called the "incomplete data", and Z the "complete data".

Assuming that the "complete data" is known, the parameters of the gray level statistical distribution associated to each class can then be computed with the Maximum Likelihood (ML) estimator on each class $k = 1, ..., K$.

Based on the assumption of electronic acquisition, we model each distribution function $P_{Y_s/X_s}(y_s/e_i)$ by a Gaussian law

$$\mathcal{N}(y; \mu_i, \sigma_i) \quad = \quad \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y-\mu_i)^2}{2\sigma_i^2}\right).$$

If $Y = (Y_1, \ldots, Y_M)$ are M random variables i.i.d. according to a "single" Gaussian law $\mathcal{N}_Y(y; \mu, \sigma)$, and $y = (y_1, \ldots, y_M)$ is a realization of Y, the ML estimator of $\mu_{\text{ML}}, \sigma^2_{\text{ML}}$ for the "complete data" is given by,

$$\hat{\mu}_{\text{ML}} = \frac{1}{M}\sum_{i=1}^{M} y_i; \qquad \hat{\sigma}^2_{\text{ML}} = \frac{1}{M}\sum_{i=1}^{M}(y_i - \hat{\mu}_{\text{ML}})^2.$$

When the segmentation result is unknown (i.e, the class label of each pixel is not supposed to be known), the situation is more complex. By assuming that the X_i are independent, this problem can be viewed as the estimation of the parameters of a K-component mixture. In this case, the observed image, or the sample $y = \{y_1, \ldots, y_N\}$, is a realization of Y with distribution function,

$$P_Y(y) = \sum_{k=1}^{K} \pi_k \, P_{Y/X_i}(y/e_k, \Phi_k),$$

where the π_k are the mixing proportions ($0 \leq \pi_k \leq 1$ for all $k = 1, \ldots, K$ and $\sum_k \pi_k = 1$). In order to obtain a reliable estimation of the parameter, we resort to the ICE algorithm. This procedure, described in detail in [3], is briefly recalled here. This method relies on an estimator $\hat{\Phi}(X, Y)$ with good asymptotic properties (like the ML estimator) for the complete data case. When X is unobserved, this procedure starts from an initial parameter vector $\Phi^{[0]}$ (not too far from the optimal one) and generates a sequence of parameter vectors leading to the optimal parameters. To this end, $\Phi^{[p+1]}$ at step $(p+1)$ is chosen as the conditional expectation of $\hat{\Phi}$ given $Y = y$, computed according to the current value $\Phi^{[p]}$. It is the best approximation of Φ in terms of the mean squares error [3]. By denoting E_p, the expectation relative to parameter vector $\Phi^{[p]}$, $\Phi^{[p+1]}$ is computed from $\Phi^{[p]}$ and $Y = y$ by $\Phi^{[p+1]} = E_p\big[\hat{\Phi}(X, Y)|Y = y\big]$. The computation of this expectation is impossible in practice, but we can approach it thanks to the law of large numbers by,

$$\Phi^{[p+1]} = \frac{1}{n}\,[\hat{\Phi}(x_{(1)}, y) + \cdots + \hat{\Phi}(x_{(n)}, y)],$$

where $x_{(i)},\ i = 1, \ldots, n$ are realizations of X drawn according to the posterior distribution $P_{X/Y}(x/y, \Phi^{[p]})$. In order to decrease the computational load, we can take $n = 1$ without altering the quality of the estimation [4]. Finally, we can use the Gibbs sampler algorithm [5] to simulate realizations of X according to the posterior distribution. For the local *a priori* model of the Gibbs sampler, we adopt a two-dimensional isotropic Potts model with a first order neighborhood [6]. In this model, there are three parameters, called "the clique parameters" denoted $\beta_1, \beta_2, \beta_3$ and associated to the horizontal, vertical, and transverse binary cliques respectively. Given this *a priori* model, the prior distribution $P_X(x)$ can be written as,

$$P_X(x) = \exp\Big\{-\sum_{<s,t>} \beta_{st}\big(1 - \delta(x_s, x_t)\big)\Big\},$$

where summation is taken over all pairs of neighboring sites and $\delta(.)$ is the Kronecker delta function. In order to favor homogeneous regions with no privileged orientation in the Gibbs sampler simulation process, we choose $\beta_{st} = \beta_1 = \beta_2 = \beta_3 = 1$. Finally, the distribution mixture parameter estimation procedure for the "incomplete data" using the ICE procedure is outlined as follows :

Parameter initialization : we can use the K-means algorithm described in [7] or an initial guess for $\Phi^{[0]}$ (not "too far" from the optimal one). Then $\Phi^{[p+1]}$ is computed from $\Phi^{[p]}$ in the following way :

1. Stochastic step : using the Gibbs sampler, one realization x is simulated according to the posterior distribution $P_{X/Y}(x/y)$, with parameter vector $\Phi^{[p]}$.

2. Estimation step : the parameter vector $\Phi^{[p+1]}$ is estimated with the ML estimator of the "complete data" corresponding to each class.

3. Repeat until convergence is achieved ; i.e., if $\Phi^{[p+1]} \not\approx \Phi^{[p]}$, for some $k = 1, \ldots, K$, return to step 1.

Based on the estimates given by the ICE procedure, we can compute an unsupervised Markov segmentation of the image. In this framework, the Markov segmentation can be viewed as a statistical labeling problem according to a global Bayesian formulation in which the distribution $P_{X/Y}(x/y) \propto \exp(-U(x, y))$ has to be maximized [6]. The corresponding posterior energy is,

$$U(x, y) = \underbrace{\sum_{s \in S} -\ln P_{Y_s/X_s}(y_s/x_s)}_{U_1(x,y)} + \underbrace{\sum_{<s,t>} \beta_{st}\Big(1 - \delta(x_s, x_t)\Big)}_{U_2(x)}$$

where U_1 expresses the adequacy between observations and labels, and U_2 represents the energy of the *a priori* model. We use the deterministic algorithm ICM [6] to minimize this global energy function. For the initialization of this algorithm, we exploit the segmentation map obtained by a ML segmentation.

3. SEGMENTATION OF THE IMAGE INTO CONTOURS

We use the same framework as in the previous section but with different distributions. Moreover, the observable data $Y = \{Y_s, s \in S\}$ now represents the norm of the gradient of the gray level in the image and each X_s is in $\{e_1 = \text{"off"}, e_2 = \text{"on"}\}$.

• First, based on empirical results, we model $P_{Y_s/X_s}(y_s/e_1)$ by a Weibull law [8],

$$\mathcal{W}(y; \min, C, \alpha) = \frac{C}{\alpha}\left(\frac{y - \min}{\alpha}\right)^{C-1} \exp\left(-\frac{(y - \min)^C}{\alpha^C}\right),$$

with $y > \min$. If $Y = (Y_1, \ldots, Y_M)$ are M random variables i.i.d. according to a "single" Weibull law $\mathcal{W}_Y(y; \min, C, \alpha)$, and $y = (y_1, \ldots, y_M)$ is a realization of Y, the ML estimator of $C_{\text{ML}}, \alpha_{\text{ML}}$ for the "complete data" is given by [9],

$$\hat{C}_{\text{ML}} = F(\hat{C}_{\text{ML}}),$$
$$\hat{\alpha}_{\text{ML}} = \left(\frac{1}{M}\sum_{i=1}^{M} \tilde{y}_i^{\hat{C}_{\text{ML}}}\right)^{\frac{1}{\hat{C}_{\text{ML}}}},$$

where $\tilde{y} = (y - \min)$ and

$$F(x) = \frac{M \sum_{i=1}^{M} \tilde{y}_i^x}{M \sum_{i=1}^{M} (\tilde{y}_i^x \ln \tilde{y}_i) - \sum_{i=1}^{M} \ln \tilde{y}_i^x \sum_{i=1}^{M} \tilde{y}_i^x}.$$

Here, we simply take $\min = -10^{-5}\nu$, where ν is the maximum of the norm of the gradient in the given image. We use an iterative method [9] to find $\hat{C}_{ML}$.

• Based on empirical results, we model the conditional density function $P_{Y_s/X_s}(y_s/e_2)$ for contour points by a mixture of three Gaussian laws,

$$\mathcal{M}(y; w_i, \mu_i, \sigma_i) = \sum_{i=1}^{3} w_i \mathcal{N}(y; \mu_i, \sigma_i).$$

The ML estimator of the parameters cannot be computed directly. We used the SEM algorithm [10] with good results in all our tests.

This algorithm is a stochastic version of the EM algorithm [11] and is identical to the ICE algorithm except for the prior distribution which is replaced by the proportion of each class. We use the ML estimator on the result given by a K-means clustering segmentation to obtain an initial solution for the parameters of this mixture.

Finally, for the local *a priori* model of the Gibbs sampler, we consider the set T formed by the contours of the sub-regions obtained in the pre-segmentation step. The prior distribution $P_X(x)$ can then be written as,

$$P_X(x) \;=\; \exp\Big\{-\sum_{<s,t>} \beta_{st}\big(1-\delta(x_s,x_t)\big)\Big\} \prod_{s\notin T} \delta(x_s,e_1)$$

where $\beta_{s,t} = 0$ whenever s or $t \notin T$, and $\beta_{s,t} = 1$ otherwise. This amounts to imposing the constraint that $V(x)$ be minimal, where

$$V(x) \;=\; \sum_{s\in T}\Big(1-\delta(x_s,e_1)\Big),$$

on the Gibbs field,

$$U(x,y) \;=\; \sum_{s\in S} -\ln P_{Y_s/X_s}(y_s/x_s) + \sum_{<s,t>} \beta_{s,t}\Big(1-\delta(x_s,x_t)\Big).$$

The constrained stochastic relaxation and constrained simulated annealing have been developed in this context [12] and could be used in the ICE procedure and segmentation step, respectively. However, we have imposed the constraint directly at all steps of the ICE procedure and ICM algorithm by setting $x_s = e_1$ whenever $s \notin T$. In particular, we initialize the ICE procedure by the parameters estimated on the complete data obtained by a K-means clustering segmentation of T into two classes based on the norm of the gradient of the gray level. The class with smallest mean is labeled e_1.

4. EXPERIMENTAL RESULTS

The effectiveness of the proposed method for detection of contours cannot be measured easily as there is no obvious definition of contours. However, one could perform goodness-of-fit tests on the distributions of the gray level and of the norm of its gradient, though this would not be sufficient. As of now, we have simply compared visually the empirical density functions and the estimated distributions. Also, one can compare visually the results of our method with other edge-detectors. For this short paper, we simply present the images obtained in two instances (cf. Fig. 1 and 2) as well as the distributions in one instance (Fig. 3). We have also included one comparison with a detection using the Canny edge-detector. One can observe that the false contours due to the pre-segmentation into sub-regions are completely removed by the segmentation into contours.

We have tested our method on a few X-rays images without success. This is certainly due to the presence of a different noise distribution in the images. Also, images that have undergone considerable JPG compression present certain artifacts with our method. However, on all images obtained by electronic acquisition that we have tested, our method gave results comparable to the ones presented here.

In our implementation, an equalization of histogram is performed before the pre-segmentation step. Then, we re-calibrate the gray levels between 0 and 255 and apply a 3×3 Gaussian mask. Finally, we re-calibrate the norm of the gradient between 0 and 100 and proceed to the estimation step (so, $\min = -0.001$). We fixed the number of iterations to 10 for the ICE procedure, to 100 for each application of the SEM algorithm and to at most 30 for each estimation of the Weibull law. The entire procedure took 100 seconds on a PC workstation 400MHz for each of the images presented here.

5. CONCLUSION

In this paper, we have described an unsupervised segmentation method for contours which proved quite adapted for the images obtained by electronic acquisition. We have presented two statistical models for the norm of the gradient of gray level at the pixels of an image, one for contour points and one for points outside contours. We have also described a Markov model with constraint which incorporates those two statistical distributions as likelihood together with a simple *a priori* model. Our model is suitable for an ICE procedure for the estimation of the parameters and an ICM algorithm, or a simulated annealing, for the segmentation. A preliminary step proceeds to the segmentation of the image into sub-regions and uses a Markov model without constraint. The results obtained are promising. One can also include a complete statistical model that takes into account the angle formed by the gradient of the gray level and the tangent to the contour curve, as will be shown elsewhere.

6. REFERENCES

[1] D. Geman, S. Geman, C. Graffigne, and P. Dong. Boundary Detection by Constrained Optimixation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7) :609–628, July 1990.

[2] M. Pérez, A. Blake, and M. Gangnet. Jetstream : Probabilistic Contour Extraction with Particles. In *Int. Conf. on Computer Vision, ICCV' 2001*, Vancouver, Canada, July 2001.

[3] F. Salzenstein and W. Pieczynski. Parameter estimation in hidden fuzzy markov random fields and image segmentation. *Graphical Models and Image Processing*, 59(4) :205–220, 1997.

[4] B. Braathen, P. Masson, and W. Pieczynski. Global and local methods of unsupervised Bayesian segmentation of images. *GRAPHICS and VISION*, 2(1) :39–52, 1993.

[5] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6) :721–741, 1984.

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

[7] S. Banks. *Signal processing, image processing and pattern recognition*. Prentice Hall, 1990.

[8] T. S. Curry, J. E. Dowdey, and R. C. Murry. *Christensen's Physics of Diagnostic Radiology*. Lea and Febiger, 1990.

[9] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Three-class markovian segmentation of high resolution sonar images. *Computer Vision and Image Understanding*, 76(3) :191–204, December 1999.

[10] G. Celeux and J. Diebolt. L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélange de densités. *Revue de statistiques appliquées*, 34(2) :35–52, 1986.

[11] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Royal Statistical Society*, pages 1–38, 1976.

[12] D. Geman and S. Geman. Relaxation and Annealing with Constraints. *Division Appl. Math., Brown Univ., Complex Systems Tech. Rep. 35*, 1987.

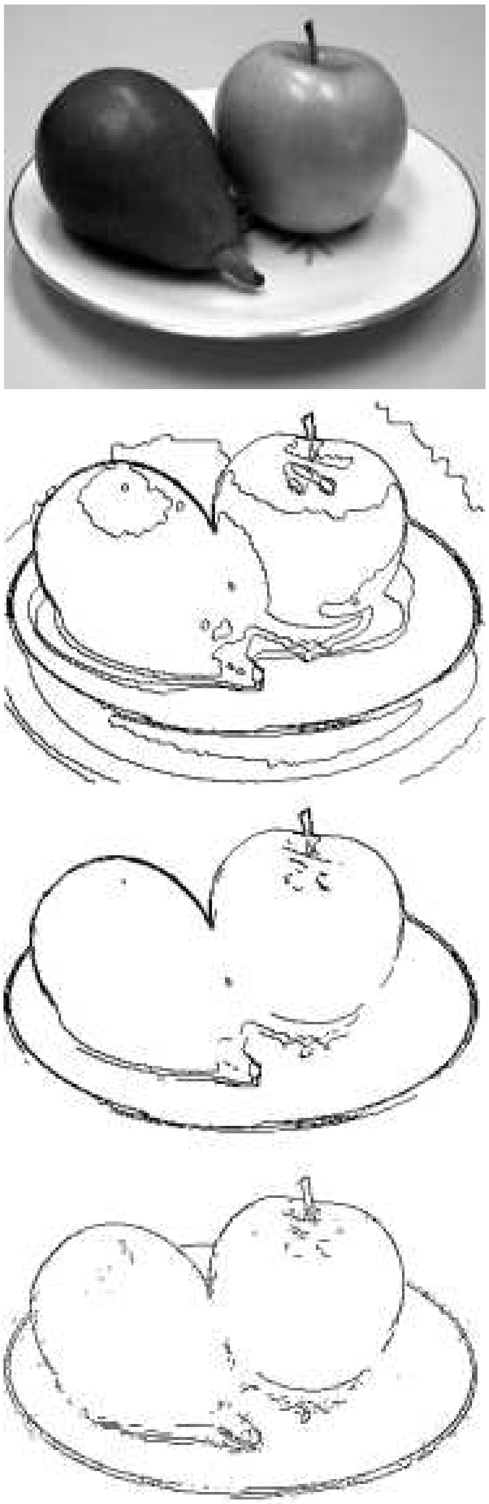

FIG. 1 – *Example of a detection of contours using the ICM and based on the parameters estimated by the ICE procedure. From Top to Bottom : original image ; contours in the pre-segmentation of the image into sub-regions ; contours detected in the image ; contours detected with the Canny algorithm with thresholds 16.25 and 25.0.*

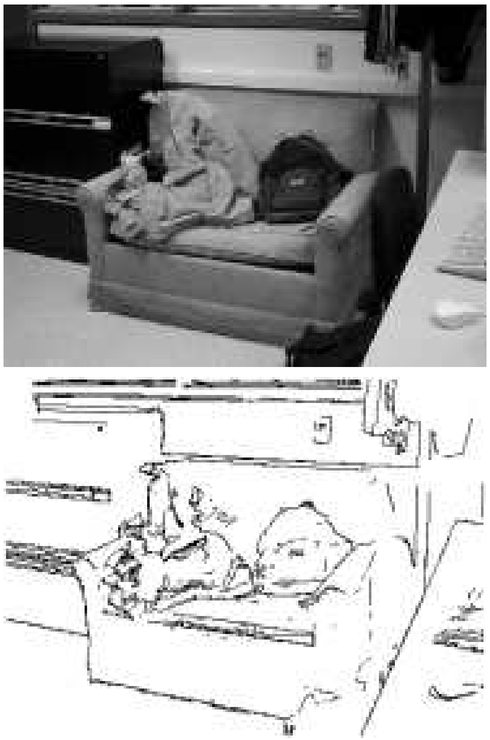

FIG. 2 – *Example of an unsupervised detection of contours using the ICM deterministic relaxation technique and based on the parameters estimated by the ICE procedure. Top : original image. Bottom : contours detected in the image.*

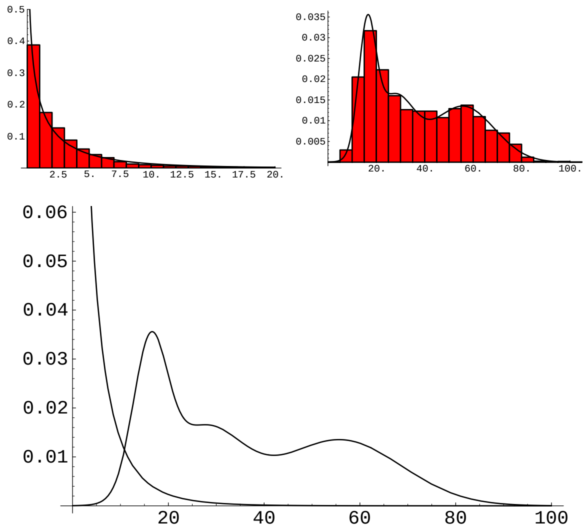

FIG. 3 – *Example of the empirical density functions of the norm of the gradient of the gray level and the distributions estimated by the ICE procedure (for the first image). Top Left : off contours. Top Right : on contours. Bottom : comparison of the two distributions.*