Empirical Bayesian Motion Segmentation

Nuno Vasconcelos, Andrew Lippman

***Abstract*— We introduce an empirical Bayesian procedure for the simultaneous segmentation of an observed motion field and estimation of the hyper-parameters of a Markov random field prior. The new approach approach exhibits the Bayesian appeal of incorporating prior beliefs, but requires only a *qualitative* description of the prior, avoiding the requirement for a *quantitative* specification of its parameters. This eliminates the need for trial-and-error strategies for the determination of these parameters and leads to better segmentations.**

Index Terms: motion segmentation, layered representations, empirical Bayesian procedures, estimation of hyper-parameters, statistical learning, expectation-maximization.

I. INTRODUCTION

The problem of motion representation is closely related to that of scene segmentation, and efficient motion estimation solutions must be capable of jointly addressing the two components. This observation has led to a generation of algorithms which iterate between optic flow estimation and segmentation [6], [9]. From a statistical perspective, such algorithms can be seen as variations of the *expectation-maximization* (EM) algorithm [3]. EM-based approaches have various attractives for segmentation, such as proceeding by taking non-greedy *soft decisions* with regards to the assignment of pixels to regions, or allowing the use of sophisticated priors, such as Markov random fields (MRFs), capable of imposing *spatial coherence* on the segmentation [8], [10], [11]. The main difficulty is, however, that such priors typically have parameters whose values are difficult to determine a priori. In practice, these parameters are commonly set to arbitrary values or adapted to the observed data through heuristics.

In this work [8], we exploit the fact that EM is itself suited for *empirical Bayesian* (EB) [2] data analysis to develop a framework for estimating the prior parameters that best explain the observed data. This eliminates the need for trial-and-error strategies for parameter setting and leads to better segmentations in less iterations.

II. BAYESIAN AND EMPIRICAL BAYESIAN INFERENCE

Assume an observer making inferences about the world property Ω, given the image feature ω. Under the Bayesian philosophy, properties in the world are random variables characterized by probability densities that express the observer's belief in their possible configurations. All inferences are based on the *posteriori* density

$$P(\Omega|\omega) = \frac{P(\omega|\Omega)P(\Omega|\gamma)}{\int P(\omega|\Omega)P(\Omega|\gamma)d\Omega}, \quad (1)$$

Nuno Vasconcelos is with Compaq Computer Corporation Cambridge Research Laboratory. He was with the MIT Media Laboratory when this work was developed. Andrew Lippman is with the MIT Media Laboratory.

where γ is a parameter that controls the shape of the property's prior.

Since observation of the data merely re-scales prior beliefs [5] it is important to get the priors right, a task which is generally difficult in practice. Typically, one does not have absolute certainty about the shape of the prior or how to set its parameters, which must be therefore regarded as random variables. That is, instead of (1) inferences should be based on

$$P(\Omega|\omega) = \frac{\int P(\omega|\Omega)P(\Omega|\gamma)P(\gamma)d\gamma}{\int\int P(\omega|\Omega)P(\Omega|\gamma)P(\gamma)d\Omega d\gamma}. \quad (2)$$

While from a perceptual standpoint such a hierarchical structure has the appeal of modeling changes of prior belief according to context (different contexts lead to different values of γ), from a computational standpoint it significantly increases the complexity of the problem. After all, the parameters of $P(\gamma)$ are themselves random variables, as well as the parameters of their densities, and so on. We are therefore caught on a endless chain of conditional probabilities which is computationally intractable.

The solution suggested by the EB philosophy is to replace γ by the estimate $\hat{\gamma}$ that maximizes the marginal distribution $P(\omega|\gamma)$. Inferences are then based on (1) using this estimate. While, strictly speaking, this approach violates the fundamental Bayesian principle that priors should not be estimated from data, in practice it leads to more sensible solutions than setting priors arbitrarily, or using priors whose main justification comes from computational simplicity. More importantly, it breaks the infinite chain of probabilities mentioned above, while still allowing context-dependent priors.

Because prior parameters are related to *observed* image features by *hidden* world properties, $P(\omega|\gamma) = \int P(\omega|\Omega)P(\Omega|\gamma)d\Omega$, the maximization of $P(\omega|\gamma)$ fits naturally into an EM framework. Hence, the EB perspective not only supports the recent trend towards the use of EM for segmentation, but extends it by providing a meaningful way to *tune* the priors to the observed data.

III. DOUBLY STOCHASTIC MOTION MODEL

Our approach to image segmentation is based on linear parametric motion models, according to which the motion of a given image region is described by $\mathbf{p}(\mathbf{x}) = \mathbf{\Psi}(\mathbf{x})\,\boldsymbol{\phi}$, where $\mathbf{x} = (x, y)^T$ is the vector of pixel coordinates in the image plane, $\mathbf{p}(\mathbf{x}) = (p_x(\mathbf{x}), p_y(\mathbf{x}))^T$ the pixel's motion, and $\boldsymbol{\phi} = (a_1, \ldots, a_P)^T$ the parameter vector which characterizes the motion of the entire region. This motion model is embedded in a probabilistic framework, where pixels are associated with classes that have a one-to-one relationship with the objects in the scene. We assume that, conditional on image $\mathbf{I}_{t-1}$ and the class of pixel $\mathbf{x}$ in image $\mathbf{I}_t$, this pixel is drawn from an independent identically distributed (iid) Gaussian process

$$P(\mathbf{I}_t(\mathbf{x})|\mathbf{z}(\mathbf{x}) = \mathbf{e}_i, \boldsymbol{\phi}_i, \mathbf{I}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{[\mathbf{I}_t(\mathbf{x}) - \mathbf{I}_{t-1}(\mathbf{x}-\mathbf{p}_i(\mathbf{x}))]^2}{2\sigma_i^2}}, \tag{3}$$

where $\mathbf{z}(\mathbf{x}) = (z_1(\mathbf{x}), \ldots, z_R(\mathbf{x}))^T$ is a vector of binary indicator variables with $\mathbf{z}(\mathbf{x}) = \mathbf{e}_i$ (where $\mathbf{e}_i$ is the i^{th} vector of the standard unitary basis) if and only if pixel $\mathbf{x}$ belongs to region i, $\mathbf{p}_i(\mathbf{x})$ the region's motion, σ_i its variance, and R the total number of regions[1]. In this work, we consider the case of affine motion where $P = 6$, $\mathbf{\Psi}(\mathbf{x})$ is a 2×6 matrix with rows $(1, x, y, 0, 0, 0)$ and $(0, 0, 0, 1, x, y)$, but the framework is generic.

Denoting by $\pi_i(\mathbf{x})$ the conditional probabilities $P(\mathbf{z}(\mathbf{x}) = \mathbf{e}_i|\mathbf{z}, \boldsymbol{\theta})$, the dependencies between the states of adjacent pixels in the images are modeled by an MRF prior

$$\pi_i(\mathbf{x}) = P(\mathbf{z}(\mathbf{x}) = \mathbf{e}_i|\mathbf{z}_\eta(\mathbf{x}), \boldsymbol{\theta}) = \frac{1}{Z_\mathbf{x}} e^{U_\mathbf{x}(\mathbf{e}_i|\boldsymbol{\theta})} \tag{4}$$

where

$$Z_\mathbf{x} = \sum_k e^{U_\mathbf{x}(\mathbf{e}_k|\boldsymbol{\theta})}, \tag{5}$$

$$U_\mathbf{x}(\mathbf{z}(\mathbf{x})|\boldsymbol{\theta}) = \sum_{C:\mathbf{x}\in C} V_C(\mathbf{z}(\mathbf{x}), \mathbf{z}_\eta(\mathbf{x})|\boldsymbol{\theta}), \tag{6}$$

$C \in \mathcal{C}$, $\mathcal{C}$ is the set of cliques in a neighborhood system $\mathcal{G}$, $\mathbf{z}_\eta(\mathbf{x})$ is the configuration of the neighbors of site $\mathbf{x}$ under $\mathcal{G}$, and $V_C(\mathbf{z}(\mathbf{x}), \mathbf{z}_\eta(\mathbf{x})|\boldsymbol{\theta})$ a function of the cliques involving site $\mathbf{x}$ and its neighbors.

While all the results can be extended to any valid $\mathcal{G}$ and V_C, we concentrate on a second-order system where the neighborhood of pixel $\mathbf{x}$, $\eta(\mathbf{x})$, consists of its 8 adjacent pixels, and

$$U_\mathbf{x}(\mathbf{z}(\mathbf{x})|\boldsymbol{\theta}) = \sum_i \left[\alpha_i z_i(\mathbf{x}) + \beta \sum_{\mathbf{y}\in\eta(\mathbf{x})} z_i(\mathbf{x})\gamma_i(\mathbf{y}) \right], \tag{7}$$

where $\gamma_i(\mathbf{x}) = E[z_i(\mathbf{x})|\theta] = P(z_i(\mathbf{x}) = 1|\theta)$. Hence, $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_R, \beta)^T$ and

$$\pi_i(\mathbf{x}) = \frac{1}{Z_\mathbf{x}} \exp\left[\alpha_i + \beta\kappa_i(\mathbf{x})\right], \tag{8}$$

where $\kappa_i(\mathbf{x}) = \sum_{\mathbf{y}\in\eta(\mathbf{x})} \gamma_i(\mathbf{y})$ is the expected number of neighbors of site $\mathbf{x}$ in state i under γ_i, β controls the degree of clustering, i.e. the likelihood of more or less class transitions between neighboring pixels, and α_i the likelihood of each of the regions.

IV. EM-BASED PARAMETER ESTIMATION

The fundamental computational problem posed by the EB framework is that of maximizing the marginal likelihood of the observed motion field as a function of the motion and MRF parameters

$$P(\mathbf{I}_t|\mathbf{I}_{t-1}, \mathbf{\Phi}, \boldsymbol{\theta}) = \sum_\mathbf{z} P(\mathbf{I}_t|\mathbf{z}, \mathbf{I}_{t-1}, \mathbf{\Phi}) P(\mathbf{z}|\mathbf{I}_{t-1}, \boldsymbol{\theta}),$$

where the summation is over all possible configurations of the hidden assignment variables vector $\mathbf{z}$, and $\mathbf{\Phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_R)^T$ is the vector of all motion parameters. The pair $(\mathbf{I}_t, \mathbf{z})$ is usually referred to as the *complete data* and has log-likelihood

$$\begin{aligned} l_c &= \log P(\mathbf{I}_t, \mathbf{z}|\mathbf{I}_{t-1}, \mathbf{\Phi}, \boldsymbol{\theta}) \\ &= \sum_\mathbf{x} \log P(\mathbf{I}_t(\mathbf{x})|\mathbf{z}, \mathbf{I}_{t-1}, \mathbf{\Phi}) + \log P(\mathbf{z}|\mathbf{I}_{t-1}, \boldsymbol{\theta}) \\ &= \sum_{\mathbf{x},i} z_i(\mathbf{x}) \log\left[P(\mathbf{I}_t(\mathbf{x})|z_i(\mathbf{x}) = 1, \mathbf{\Phi})\right] + \log P(\mathbf{z}|\boldsymbol{\theta}), \end{aligned} \tag{9}$$

where, for simplicity, we have dropped the dependence on $\mathbf{I}_{t-1}$.

The EM algorithm maximizes the likelihood of the incomplete, observed, data by iterating between two steps that act on the log-likelihood of the complete data. The E-step computes the so-called Q function

$$Q(\mathbf{\Xi}'|\mathbf{\Xi}^{(p)}) = E[l_c|\mathbf{I}_t, \mathbf{\Xi}^{(p)}] \tag{10}$$

where $\mathbf{\Xi}^{(p)} = (\mathbf{\Phi}^{(p)}, \boldsymbol{\theta}^{(p)})^T$ is the vector of parameter estimates obtained in the previous iteration. The M-step then maximizes this function with respect to $\mathbf{\Phi}$ and $\boldsymbol{\theta}$. In general, both steps are analytically intractable.

A. Segmentation with known prior parameters

The problem can be significantly simplified by assuming that the prior parameters are known. In this case, there are no parameters to estimate in $E[\log P(\mathbf{z}|\boldsymbol{\theta})|\mathbf{I}_t, \mathbf{\Xi}^{(p)}]$ and this term can be eliminated. Nevertheless, the exact computation of the remaining $E[z_i(\mathbf{x})|\mathbf{I}_t, \mathbf{\Xi}^{(p)}]$ is still a tremendous challenge, which can only be addressed through Markov chain Monte Carlo procedures. However, nesting such procedures inside the EM iteration would lead to a prohibitive amount of computation.

Zhang et al. [11] have shown that if Besag's *pseudo-likelihood* (PL) approximation [1]

$$P(\mathbf{z}|\boldsymbol{\theta}) \approx \prod_\mathbf{x} P(\mathbf{z}(\mathbf{x})|\mathbf{z}_\eta(\mathbf{x}), \boldsymbol{\theta}) \tag{11}$$

is used, then it is reasonable to assume that $\gamma_i(\mathbf{x}) \approx \pi_i(\mathbf{x})$ and, from Bayes rule,

$$\begin{aligned} \lambda_i(\mathbf{x}) &= E[z_i(\mathbf{x})|\mathbf{I}_t, \mathbf{\Xi}] = P(z_i(\mathbf{x}) = 1|\mathbf{I}_t, \mathbf{\Xi}) \\ &\approx \frac{P(\mathbf{I}_t(\mathbf{x})|z_i(\mathbf{x}) = 1, \mathbf{\Phi})\,\pi_i(\mathbf{x})}{\sum_k P(\mathbf{I}_t(\mathbf{x})|z_k(\mathbf{x}) = 1, \mathbf{\Phi})\,\pi_k(\mathbf{x})}. \end{aligned} \tag{12}$$

This suggests an iterative procedure for the computation of the expectations required by the E-step: at iteration p, 1) compute the conditional assignment probabilities, $\pi_i^{(p)}(\mathbf{x}), i = 1, \ldots, R$, sequentially by visiting each of the pixels in the image in a pre-defined (typically raster scan) order, assuming at each $\mathbf{x}$ that the current posterior estimates $\lambda_i^{(p-1)}(\mathbf{x})$ of the assignment probabilities of the neighboring pixels are the true marginals $\gamma_i(\mathbf{x})$, i.e.

$$\pi_i^{(p)}(\mathbf{x}) = \frac{1}{Z_\mathbf{x}} \exp\left[\alpha_i + \beta \sum_{\mathbf{y}\in\eta(\mathbf{x})} \lambda_i^{(p-1)}(\mathbf{y})\right], \tag{13}$$

and 2) update the posterior assignment probabilities using (12). This is an extension of Besag's *iterated conditional modes*

[1] Assumed to be known.

(ICM) procedure, capable of supporting the soft decisions required by EM, and we will therefore refer to it as *iterated conditional probabilities* (ICP). It was first proposed, in the context of texture segmentation, by Zhang et al. [11].

B. Empirical Bayesian Segmentation

The procedure above has two major limitations. First, the assumption that the prior parameters are known is, in general, unrealistic. Second, with arbitrary parameter selections, there is no way to guarantee that the posterior estimates $\lambda_i^{(p)}(\mathbf{x})$ converge to the true marginals $\gamma_i(\mathbf{x})$ and it is difficult to justify the assumption behind step 1). We now show that 1) EB estimates provide a natural answer to these two problems and 2) under the PL and ICP approximations these estimates do not imply a significant increase in complexity: the only additional requirement is a simple concave maximization in the M-step, no extra computation being required by the E-step.

1) The E-step: The evaluation of $E[\log P(\mathbf{z}|\boldsymbol{\theta})|\mathbf{I}_t, \Xi^{(p)})]$ is significantly simplified by the PL approximation since, from (11), (4), and the binary nature of $z_i(\mathbf{x})$,

$$E[\log P(\mathbf{z}|\boldsymbol{\theta})|\mathbf{I}_t, \Xi^{(p)}] \approx \sum_{\mathbf{x},i} E[z_i(\mathbf{x}) \log \pi_i(\mathbf{x})|\mathbf{I}_t, \Xi^{(p)}].$$

Furthermore, under ICP, whenever a pixel is visited $\pi_i(\mathbf{x})$ is approximated by $\pi_i^{(p)}(\mathbf{x})$ which, as shown by (13), has no random components. Hence,

$$E[\log P(\mathbf{z}|\boldsymbol{\theta})|\mathbf{I}_t, \Xi^{(p)}] \approx \sum_{\mathbf{x},i} E[z_i(\mathbf{x})|\mathbf{I}_t, \Xi^{(p)}] \log \pi_i^{(p)}(\mathbf{x}) \tag{14}$$

i.e. $E[\log P(\mathbf{z}|\boldsymbol{\theta})|\mathbf{I}_t, \Xi^{(p)}]$ simply requires the evaluation of the expectations $\lambda_i^{(p)}$, which were already necessary for the evaluation of the remaining components of the Q function. Therefore, under the above approximations, there is no computational cost in evaluating Q completely.

2) The M-step: The M-step maximizes the Q function with respect to both the motion and MRF parameters. Combining (10), (9), (12), (14) and (3)

$$Q(\mathbf{\Phi}'|\mathbf{\Phi}^{(p)}) = \sum_{\mathbf{x},i} \lambda_i^{(p)}(\mathbf{x}) \left\{ \log \pi_i(\mathbf{x}) - \frac{1}{2}\log(2\pi\sigma_i^2) - \frac{[\mathbf{I}_t(\mathbf{x}) - \mathbf{I}_{t-1}(\mathbf{x}-\mathbf{p}_i(\mathbf{x}))]^2}{2\sigma_i^2} \right\}.$$

The maximization of Q with respect to the motion parameters is a variation of the least-squares problem found in image registration [9], and solvable by any of the standard techniques from the registration literature. In our implementation, we use Gauss-Newton's method. The maximization with respect to the MRF parameters depends only on the first term. Noticing, from (4) and (5) that

$$\nabla_\theta \log \pi_i(\mathbf{x}) = \nabla_\theta U_{\mathbf{x}}(\mathbf{e}_i) - \sum_k \nabla_\theta U_{\mathbf{x}}(\mathbf{e}_k)\pi_k(\mathbf{x}),$$

where for simplicity we have omitted the dependence of $U_{\mathbf{x}}$ on θ, it follows that

$$\begin{aligned}\nabla_\theta Q &= \sum_{\mathbf{x},i} \lambda_i(\mathbf{x}) \nabla_\theta U_{\mathbf{x}}(\mathbf{e}_i) - \sum_{\mathbf{x},i} \lambda_i(\mathbf{x}) \sum_k \nabla_\theta U_{\mathbf{x}}(\mathbf{e}_k)\pi_k(\mathbf{x}),\\ &= \sum_{\mathbf{x}} \left(E[\nabla_\theta U_{\mathbf{x}}(\mathbf{z}(\mathbf{x}))]_{post} - E[\nabla_\theta U_{\mathbf{x}}(\mathbf{z}(\mathbf{x}))]_{prior}\right),\end{aligned}$$

where we have used the fact that $\sum_i \lambda_i(\mathbf{x}) = 1$ and subscripts $prior$ and $post$ indicate expectations taken over the prior (π_i) and posterior (λ_i) distributions, respectively. Similarly,

$$\begin{aligned}\nabla_\theta^2 Q &= \sum_{\mathbf{x}} E[\nabla_\theta^2 U_{\mathbf{x}}(\mathbf{z}(\mathbf{x}))]_{post} - \sum_{\mathbf{x}} E[\nabla_\theta^2 U_{\mathbf{x}}(\mathbf{z}(\mathbf{x}))]_{prior}\\ &- \sum_{\mathbf{x}} Cov[\nabla_\theta U_{\mathbf{x}}(\mathbf{z}(\mathbf{x}))]_{prior},\end{aligned}$$

where $Cov[\nabla_\theta U_{\mathbf{x}}(\mathbf{z}(\mathbf{x}))]_{prior}$ is the covariance of the gradient $\nabla_\theta U_{\mathbf{x}}(\mathbf{z}(\mathbf{x}))$ under the prior distribution. Typically, $U_{\mathbf{x}}(\mathbf{z}(\mathbf{x}), \theta)$ is a linear function of θ, and this expression reduces to the third term. Since the covariance matrixes in the summation are positive definite, $\nabla_\theta^2 Q$ is negative definite over the entire parameter space and the Q function is concave.

This implies that standard non-linear programming techniques such as Newton's method will achieve its global maximum in few iterations and, together with the fact that no extra computation is required in the E-step, makes the cost of EB segmentation only marginally superior to that required when the parameters are pre-set. In our implementation, we have indeed used Newton's method to solve the maximization with respect to the hyper-parameters.

For the specific potentials of (7), we have

$$\frac{\partial}{\partial \alpha_i} U_{\mathbf{x}}(\mathbf{e}_i) = 1 \quad \text{and} \quad \frac{\partial}{\partial \beta} U_{\mathbf{x}}(\mathbf{e}_i) = \kappa_i(\mathbf{x}),$$

from which

$$\begin{aligned}\frac{\partial}{\partial \alpha_i} Q &= \sum_{\mathbf{x}} \left[\lambda_i^{(p)}(\mathbf{x}) - \pi_i^{(p)}(\mathbf{x})\right]\\ \frac{\partial}{\partial \beta} Q &= \sum_{\mathbf{x}} E[\kappa(\mathbf{x})]_{post} - E[\kappa(\mathbf{x})]_{prior}.\end{aligned}$$

Hence, a step in the direction of the gradient changes α so that, at each pixel, the prior assignment probabilities move towards the posterior assignment probabilities derived from the observed motion. Similarly, a gradient step changes β so that, at each pixel, the expected number of neighbors in the same state as the pixel is equal under both the prior and the posterior distributions. I.e. EB estimation sets the hyper-parameters to the values that best explain the observed data, both in terms of assignment probabilities and average number of neighbors in the same state as the neighborhood's central pixel. This not only is intuitive, but justifies the assumption behind step 1) of section IV-A.

There remains, however, one problem: a pixel whose motion is poorly explained by all the models in $\mathbf{\Phi}^{(p)}$ will originate zero class-conditional likelihoods and the corresponding posterior region assignment probabilities $\lambda_i(\mathbf{x})$ will be undefined. To avoid this problem, we rely on the fact that a pixel which cannot be explained by any of the models is an outlier, and set the corresponding $\lambda_i(\mathbf{x})$ to zero for all i. This is equivalent to assuming a background outlier process of uniform likelihood over the entire parameter space, and originates robust estimates without increasing the complexity of the M-step [4].

V. Experimental results and conclusions

We tested EB motion segmentation on various video sequences, starting with a synthetic sequence that allows objective performance evaluation. The sequence is a realization

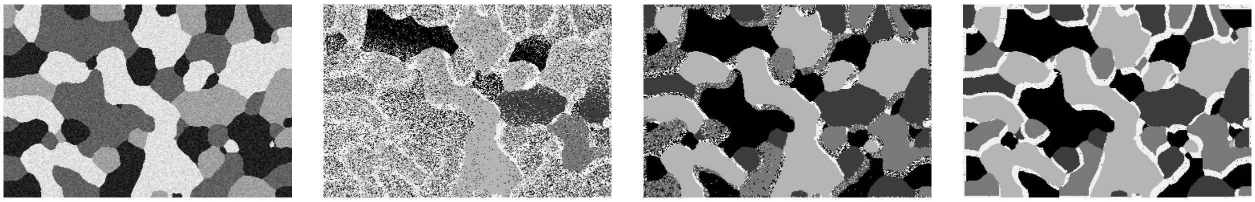
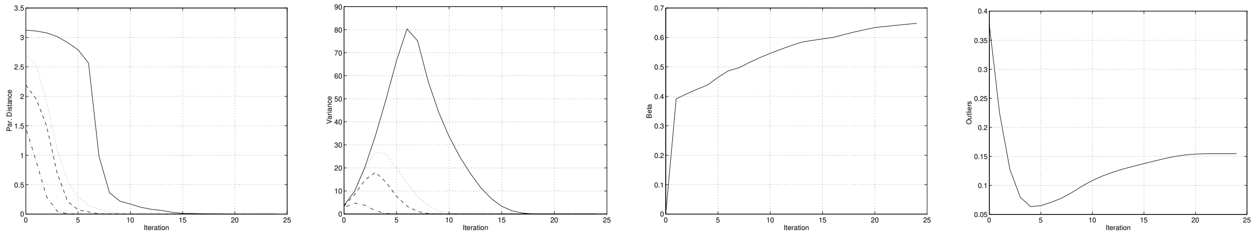


Fig. 1. Segmentation of a synthetic sequence. From left to right: first frame, segmentation after 1, 10, and 20 EM iterations.



a) b) c) d)

Fig. 2. Evolution of several parameters of the motion model as a function of the EM iteration: a) distance between parameter estimates and true values, b) variance of the Gaussian associated with each mixture component, c) clustering parameter β, and d) percentage of pixels classified as outliers. In a) and b) each curve corresponds to one component of the mixture model.

Fig. 3. Segmentations obtained for three sequences with diverse characteristics. Top: first frame of each sequence, bottom: segmentation.

Fig. 4. *Flower garden* segmentations with the MRF clustering parameter set to arbitrary values (left to right $\beta = 0$, $\beta = 0.7$, and $\beta = 1.2$).

of the model of section III: a segmentation mask was first drawn, using Gibbs sampling, from the Gibbs distribution (8) with parameters $\alpha_i = 0, \beta = 0.7$. A different texture (constant intensity plus additive Gaussian noise) was then assigned to each region. Finally, subsequent frames of the sequence were created according to (3) with $\sigma_i = 0$.

Figure 1 presents the segmentations obtained after 1, 10, and 20 iterations. The algorithm converges quickly to a segmentation that is visually very close to the optimal and, as expected, outlying pixels (shown in white) are located mostly along occlusion boundaries, where the assumptions behind the motion model break down. We emphasize that in all the examples the segmentations are shown exactly as they are produced by the segmentation algorithm. In fact, we have not even tried to reassign the outlying pixels (shown in white in all figures) to the segmented regions.

Figure 2 a) shows that the motion parameters converge quickly to the true values. Quite interesting is the behavior of the variance estimates, shown in b): they increase in the early iterations, but decrease and converge to zero as the segmentation converges. This variance increase, and the corresponding spread of the associated Gaussians, allows each region to accept new pixels and, therefore, progress towards the optimal estimates. Notice that each mixture component is subject to a different variance increase, whose magnitude seems to be a function of the error of the initial estimate for its motion parameters.

Yet another interesting feature, visible in c), is the fact that the β tends to be small in early iterations increasing as the segmentation moves towards convergence, and converging to a final estimate which is very close to the parameter's true value. Because lower values of β lead to a smaller constraint for a pixel to be in the same state as the pixels in its neighborhood, this behavior allows pixels to move freely between regions when there is a lot of uncertainty with regards to the segmentation (early iterations), constraining this freedom as the process approaches convergence. It is revealing to notice that heuristic procedures based on parameter updates *qualitatively* similar to the ones above have been proposed in the literature [1], [4], [7] as a way to escape local minima and to improve the speed and accuracy of segmentation algorithms. These procedures rely, however, on pre-defined updating schedules which are hard to justify, and do not generalize well across different types of imagery. On the other hand, the updates associated with the EB framework have a theoretical justification and are completely driven by the data, i.e. generic.

The MRF clustering parameter can also be viewed as the inverse of the temperature usually associated with Gibbs distributions. Under this perspective, the behavior of plot c), is that of an annealing process, where optimization is performed over a succession of prior distributions with characteristics controlled by this temperature: high temperatures lead to approximately uniform prior region assignment probabilities in the early iterations (allowing pixels to switch between regions very easily), the distribution becoming more peaked at later iterations where lower temperatures are used (making it difficult for pixels to change state). The EB approach is not a true simulated annealing procedure, as it is completely driven by the data. No pre-defined schedules are required, and convergence is fast but not necessarily to a global maximum of the likelihood of the observed data. It exhibits however a behavior *qualitatively* similar to that of simulated annealing which appears to give it some robustness against local minima.

Figure 3 presents segmentations for four real sequences with different levels of difficulty (from pure affine motion and small occlusion on the left, to motion that can only be coarsely approximated by an affine model and significant amounts unveiled background on the right). The advantages of the EB approach are illustrated by Figure 4, where we present three segmentations of "Flower garden" obtained by setting the MRF parameters to arbitrary values. The figure shows that, in this case, the segmentation depends critically on the choice of the clustering parameter β. While small values of clustering lead to very noisy segmentations (left), large values originate segmentations with reduced accuracy near region boundaries (right). And even though it may be possible to obtain good results by a trial-and-error strategy, we were not able to obtain, in this way, a segmentation as good as or better than the EB one. For example, setting β to the value of the optimal EB estimate throughout all iterations leads to the segmentation shown in the center, which is still significantly less accurate than the one in Figure 3 (notice the leakage between the house and sky, and house and flower bed regions, and between the areas of tree detail and sky).

References

[1] J. Besag. On the Statistic Analysis of Dirty Pictures. *J. R. Statistical Society B*, 48(3):259–302, 1986.

[2] B. Carlin and T. Louis. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman Hall, 1996.

[3] A. Dempster, N. Laird, and D. Rubin. Maximum-likelihood from Incomplete Data via the EM Algorithm. *J. of the Royal Statistical Society*, B-39, 1977.

[4] A. Jepson and M. Black. Mixture Models for Image Representation. Technical Report ARK96-PUB-54, University of Toronto, March 1996.

[5] A. Jepson, W. Richards, and D. Knill. Modal Structure and Reliable Inference. In D. Knill and W. Richars, editors, *Perception as Bayesian Inference*. Cambridge Univ. Press, 1996.

[6] D. Murray and B. Buxton. Scene Segmentation from Visual Motion Using Global Optimization. *IEEE Trans. on Pattern. Analysis and Machine Intelligence*, Vol. PAMI-9, March 1987.

[7] T. Pappas. An Adaptive Clustering Algorithm for Image Segmentation. *IEEE Trans. on Signal Processing*, Vol. 40, April 1992.

[8] N. Vasconcelos and A. Lippman. Empirical Bayesian EM-based Motion Segmentation. In *Proc. IEEE Computer Vision and Pattern Recognition Conf.*, San Juan, Puerto Rico, 1997.

[9] J. Wang and E. Adelson. Representing Moving Images with Layers. *IEEE Trans. on Image Processing*, Vol. 3, September 1994.

[10] Y. Weiss and E. Adelson. A Unified Mixture Framework for Motion Segmentation: Incorporating Spatial Coherence and Estimating the Number of Models. In *Proc. Computer Vision and Pattern Recognition Conf.*, 1996.

[11] J. Zhang, J. Modestino, and D. Langan. Maximum-Likelihood Parameter Estimation for Unsupervised Stochastic Model-Based Image Segmentation. *IEEE Trans. on Image Processing*, Vol. 3, July 1994.