A Label Field Fusion Bayesian Model and Its Penalized Maximum Rand Estimator for Image Segmentation

Max Mignotte

***Abstract*—This paper presents a novel segmentation approach based on a Markov random field (MRF) fusion model which aims at combining several segmentation results associated with simpler clustering models in order to achieve a more reliable and accurate segmentation result. The proposed fusion model is derived from the recently introduced probabilistic Rand measure for comparing one segmentation result to one or more manual segmentations of the same image. This non-parametric measure allows us to easily derive an appealing fusion model of label fields, easily expressed as a Gibbs distribution, or as a nonstationary MRF model defined on a *complete* graph. Concretely, this Gibbs energy model encodes the set of binary constraints, in terms of pairs of pixel labels, provided by each segmentation results to be fused. Combined with a prior distribution, this energy-based Gibbs model also allows for definition of an interesting penalized maximum probabilistic rand estimator with which the fusion of simple, quickly estimated, segmentation results appears as an interesting alternative to complex segmentation models existing in the literature. This fusion framework has been successfully applied on the Berkeley image database. The experiments reported in this paper demonstrate that the proposed method is efficient in terms of visual evaluation and quantitative performance measures and performs well compared to the best existing state-of-the-art segmentation methods recently proposed in the literature.**

***Index Terms*—Bayesian model, Berkeley image database, color textured image segmentation, energy-based model, label field fusion, Markovian (MRF) model, probabilistic Rand index.**

I. Introduction

IMAGE segmentation is a frequent preprocessing step which consists of achieving a compact region-based description of the image scene by decomposing it into spatially coherent regions with similar attributes. This low-level vision task is often the preliminary and also crucial step for many image understanding algorithms and computer vision applications.

A number of methods have been proposed and studied in the last decades to solve the difficult problem of textured image segmentation. Among them, we can cite clustering algorithms [1], spatial-based segmentation methods which exploit the connectivity information between neighboring pixels and have led to Markov Random Field (MRF)-based statistical models [2], mean-shift-based techniques [3], [4], graph-based [5], [6], variational methods [7], [8], or by region-based split and merge procedures, sometimes directly expressed by a global energy function to be optimized [9].

Years of research in segmentation have demonstrated that significant improvements on the final segmentation results may be achieved either by using notably more sophisticated feature selection procedures, or more elaborate clustering techniques (sometimes involving a mixture of different or non-Gaussian distributions for the multidimensional texture features [10], [11]) or by taking into account prior distribution on the labels, region process, or the number of classes [9], [12], [13]. In all cases, these improvements lead to computationally expensive segmentation algorithms and, in the case of energy-based segmentation models, to costly optimization techniques.

The segmentation approach, proposed in this paper, is conceptually different and explores another strategy initially introduced in [14]. Instead of considering an elaborate and *better designed* segmentation model of textured natural image, our technique explores the possible alternative of fusing (i.e., efficiently combining) several quickly estimated segmentation maps associated with simpler segmentation models for a final reliable and accurate segmentation result. These initial segmentations to be fused can be given either by different algorithms or by the same algorithm with different values of the internal parameters such as several K-means clustering results with different values of K, or by several K-means results using different distance metrics, and applied on an input image possibly expressed in different color spaces or by other means.

The fusion model, presented in this paper, is derived from the recently introduced probabilistic rand index (PRI) [15], [16] which measures the agreement of one segmentation result to multiple (manually generated) ground-truth segmentations. This measure efficiently takes into account the inherent variation existing across hand-labeled possible segmentations. We will show that this non-parametric measure allows us to derive an appealing fusion model of label fields, easily expressed as a Gibbs distribution, or as a nonstationary MRF model defined on a *complete* graph. Finally, this fusion model emerges as a classical optimization problem in which the Gibbs energy function related to this model has to be minimized. In other words, or analytically expressed in the regularization framework, each *quickly estimated* segmentation (to be fused) provides a set of constraints in terms of pairs of pixel labels (i.e., binary cliques) that should be equal or not. Finally, our fusion result is found

Manuscript received February 20, 2009; revised February 06, 2010. First published March 11, 2010; current version published May 14, 2010. This work was supported by a NSERC individual research grant. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Peter C. Doerschuk.

The author is with the Département d'Informatique et de Recherche Opérationnelle (DIRO), Université de Montréal, Faculté des Arts et des Sciences, Montréal H3C 3J7 QC, Canada (e-mail: mignotte@iro.umontreal.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2010.2044965

by searching for a segmentation map that minimizes an energy function encoding this precomputed set of binary constraints (thus optimizing the so-called PRI criterion). In our application, this final optimization task is performed by a robust multiresolution coarse-to-fine minimization strategy. This fusion of simple, quickly estimated segmentation results appears as an interesting alternative to complex, computationally demanding segmentation models existing in the literature. This new strategy of segmentation is validated in the Berkeley natural image database (also containing, for quantitative evaluations, ground truth segmentations obtained from human subjects).

Conceptually, our fusion strategy is in the framework of the so-called *decision fusion* approaches recently proposed in clustering or imagery [17]–[21]. With these methods, a series of energy functions are first minimized before their outputs (i.e., their decisions) are merged. Following this strategy, Fred *et al.* [17] have explored the idea of evidence accumulation for combining the results of multiple clusterings. Reed *et al.* have proposed a Gibbs energy-based fusion model that differs from ours in the likelihood and prior energy design, as final merging procedure (for the fusion of large scale classified sonar image [21]). More precisely, Reed *et al.* employed a voting scheme-based likelihood regularized by an isotropic Markov random field priorly used to inpaint regions where the likelihood decision is not available.

More generally, the concept of combining classifiers for the improvement of the performance of individual classifiers is known, in machine learning field, as a committee machine or mixture of experts [22], [23]. In this context, Dietterich [23] have provided an accessible and informal reasoning, from statistical, computational and representational viewpoints, of why ensembles can improve results. In this recent field of research, two major categories of committee machines are generally found in the literature. Our fusion decision approach is in the category of the committee machine model that utilizes an ensemble of classifiers with a static structure type. In this class of committee machines, the responses of several classifiers are combined by means of a mechanism that does not involve the input data (contrary to the dynamic structure type-based mixture of experts). In order to create an efficient ensemble of classifiers, three major categories of methods have been suggested whose goal is to promote diversity in order to increase efficiency of the final classification result. This can be done either by using different subsets of the input data, either by using a great diversity of the behavior between classifiers on the input data or finally by using the diversity of the behavior of the input data. Conceptually, our ensemble of classifiers is in this third category, since we intend to express the input data in different color spaces, thus encouraging diversity and different properties such as data decorrelation, decoupling effects, perceptually uniform metrics, compaction and invariance to various features, etc. In this framework, the combination itself can be performed according to several strategies or criteria (e.g., weighted majority vote, probability rules: sum, product, mean, median, classifier as combiner, etc.) but, none (to our knowledge) uses the PRI fusion (PRIF) criterion.

Our segmentation strategy, based on the fusion of *quickly estimated* segmentation maps, is similar to the one proposed in [14] but the criterion which is now used in this new fusion model is different. In [14], the fusion strategy can be viewed as a two-step hierarchical segmentation procedure in which the first step remains identical and a set of initial input texton segmentation maps (in each color space) is estimated. Second, a final clustering, taking into account this mixture of textons (expressed in the set of different color space) is then used as a discriminant feature descriptor for a final K-mean clustering whose output is the final fused segmentation map. Contrary to the fusion model presented in this paper, this second step (fusion of texton segmentation maps) is thus achieved in the intra-class inertia sense which is also the so-called squared-error criterion of the K-mean algorithm.

Let us add that a conceptually different label field fusion model has been also recently introduced in [24] with the goal of blending a spatial segmentation (region map) and a quickly estimated and to-be-refined application field (e.g., motion estimation/segmentation field, occlusion map, etc.). The goal of the fusion procedure explained in [24] is to locally fuse label fields involving labels of two different natures at different level of abstraction (i.e., pixel-wise and region-wise). More precisely, its goal is to iteratively modify the application field to make its regions fit the color regions of the spatial segmentation with the assumption that the color segmentation is more detailed than the regions of the application field. In this way, misclassified pixels in the application field (false positives and false negatives) are filtered out and blobby shapes are sharpened, resulting in a more accurate final application label field.

The remainder of this paper is organized as follows. Section II describes the proposed Bayesian fusion model. Section III describes the optimization strategy used to minimize the Gibbs energy field related to this model and Section IV describes the segmentation model whose outputs will be fused by our model. Finally, Section V presents a set of experimental results and comparisons with existing segmentation techniques.

II. Proposed Fusion Model

A. Rand Index

The Rand index [25] is a clustering quality metric that measures the agreement of the clustering result with a given ground truth. This non-parametric statistical measure was recently used in image segmentation [16] as a quantitative and perceptually interesting measure to compare automatic segmentation of an image to a ground truth segmentation (e.g., a manually hand-segmented image given by an expert) and/or to objectively evaluate the efficiency of several unsupervised segmentation methods.

Let n_s be the number of pixels assigned to the same region (i.e., matched pairs) in both the segmentation to be evaluated (S^{test}) and the ground truth segmentation S^{gt}, and n_d be the number of pairs of pixels assigned to different regions (i.e., mismatched pairs) in S^{test} and S^{gt}. The Rand index is defined as the ratio of $(n_s + n_d)$ to the total number of pixel pairs, i.e., $N(N-1)/2$ for an image of size N pixels. More formally [16], if $\{l_i^{S_{\text{test}}}\}$ and $\{l_i^{S^{gt}}\}$ designate the set of region labels respectively associated to the segmentation maps S^{test} and S^{gt} at pixel location $\mathbf{x}_i$ and where $\mathcal{I}$ is an indicator function, the Rand index

is given by the following relation:

$$\mathrm{Rand}(S^{\mathrm{test}}, S^{gt}) = \frac{1}{\frac{N(N-1)}{2}} \times \sum_{i,j\ i<j} \left[\overbrace{\mathcal{I}\left(l_i^{S^{\mathrm{test}}} = l_j^{S^{\mathrm{test}}} \text{ and } l_i^{gt} = l_j^{gt}\right)}^{n_s} + \overbrace{\mathcal{I}\left(l_i^{S^{\mathrm{test}}} \neq l_j^{S^{\mathrm{test}}} \text{ and } l_i^{S^{gt}} \neq l_j^{S^{gt}}\right)}^{n_d} \right] \quad (1)$$

which simply computes the proportion (value ranging from 0 to 1) of pairs of pixels with compatible region label relationships between the two segmentations to be compared. A value of 1 indicates that the two segmentations are identical and a value of 0 indicates that the two segmentations do not agree on any pair of points (e.g., when all the pixels are gathered in a single region in one segmentation whereas the other segmentation assigns each pixel to an individual region). When the number of labels in S^{test} and S^{gt} are much smaller than the number of data points N, a computationally inexpensive estimator of the Rand index can be found in [16].

B. Probabilistic Rand Index (PRI)

The PRI was recently introduced by Unnikrishnan [16] to take into account the inherent variability of possible interpretations between human observers of an image, i.e., the multiple acceptable ground truth segmentations associated with each natural image. This variability between observers, recently highlighted by the Berkeley segmentation dataset [26] is due to the fact that each human chooses to segment an image at different levels of detail. This variability is also due image segmentation being an ill-posed problem, which exhibits multiple solutions for the different possible values of the number of classes not known *a priori*.

Hence, in the absence of a *unique* ground-truth segmentation, the clustering quality measure has to quantify the agreement of an automatic segmentation (i.e., given by an algorithm) with the variation in a set of available manual segmentations representing, in fact, a very small sample of the set of all possible perceptually consistent interpretations of an image [15]. The authors [16] address this concern by soft nonuniform weighting of pixel pairs as a means of accounting for this variability in the ground truth set. More formally, let us consider a set of L manually segmented (ground truth) images $\{S_1^{gt}, S_2^{gt}, \ldots, S_L^{gt}\}$ corresponding to an image of size N. Let S^{test} be the segmentation to be compared with the manually labeled set and $\{l_i^{S_k^{gt}}\}$ designates the set of region labels associated with the segmentation maps S_k^{gt} at pixel location $\mathbf{x}_i$, the probabilistic RI is defined by

$$\mathrm{PRand}\left(S^{\mathrm{test}}, \{S_k^{gt}\}\right) = \frac{1}{\frac{N(N-1)}{2}} \sum_{i,j\ i<j} \times \left[p_{ij}\, \mathcal{I}\left(l_i^{S^{\mathrm{test}}} = l_j^{S^{\mathrm{test}}}\right) + (1-p_{ij}) \mathcal{I}\left(l_i^{S^{\mathrm{test}}} \neq l_j^{S^{\mathrm{test}}}\right) \right] \quad (2)$$

where a good choice for the estimator of p_{ij} (the probability of the pixel i and j having the same label across $\{S_k^{gt}\}$) is simply given by the empirical proportion

$$p_{ij} = \frac{1}{L} \sum_{k=0}^{k=L} \delta\left(l_i^{S_k^{gt}}, l_j^{S_k^{gt}}\right) \quad (3)$$

where δ is the delta Kronecker function. In this way, the PRI measure is simply the mean of the Rand index computed between each pair $(S^{\mathrm{test}}, S_k^{gt})\,(k = 0, \ldots, L)$ [16]. As a consequence, the PRI measure will favor (i.e., give a high score to) a resulting acceptable segmentation map which is consistent with most of the segmentation results given by human experts. More precisely, the resulting segmentation could result in a compromise or a consensus, in terms of level of details and contour accuracy exhibited by each ground-truth segmentations. Fig. 8 gives a fusion map example, using a set of manually generated segmentations exhibiting a high variation, in terms of level of details. Let us add that this probabilistic metric is not degenerate; all the bad segmentations will give a low score without exception [16].

C. Generative Gibbs Distribution Model of Correct Segmentations

As indicated in [15], the set $\{p_{ij}\}$ (i.e., the pairwise empirical probabilities for each pixel pair computed over $\{S_k^{gt}\}$) defines an appealing generative model of *correct* segmentation ($S = \{l_i\}$) for the image, easily expressed as a Gibbs distribution. In this way, the Gibbs distribution, generative model of *correct* segmentation, which can also be considered as a likelihood of S, in the PRI sense, may be expressed as

$$P\left(\{p_{ij}\}|S = \{l_i\}\right) = \frac{1}{Z} \exp\left(\mathrm{PRand}\left(S, \{S_k^{gt}\}\right)\right) = \frac{1}{Z} \exp\left(\frac{1}{T} \sum_{\langle i,j \rangle} p_{ij} \delta(l_i, l_j) + (1-p_{ij})\left[1 - \delta(l_i, l_j)\right] \right)$$

where $\langle i, j \rangle$ is the set of second order cliques or binary cliques of a Markov random field (MRF) model defined on a *complete* graph (each node or pixel $\mathbf{x}_i$ is connected to all other pixels of the image) and $T = N(N-1)$ is the temperature factor of this Boltzmann–Gibbs distribution which is twice less than the normalization factor of the Rand Index in (1) or (2) since there are twice more binary cliques $\langle i, j \rangle$ than pairs of pixels for which $i < j$. Z is the constant partition function. After simplification, this yields

$$\begin{aligned} P\left(\{p_{ij}\}|S = \{l_i\}\right) &= \frac{1}{Z} \exp\left(\mathrm{PRand}\left(S, \{S_k^{gt}\}\right)\right) \\ &= \frac{1}{Z} \exp\left(\left[\underbrace{1 - \frac{1}{T} \sum_{\langle i,j \rangle} p_{ij}}_{\Upsilon} \right] - \frac{1}{T} \sum_{\langle i,j \rangle} \delta(l_i, l_j)[1 - 2p_{ij}] \right) \\ &= \frac{1}{Z_1} \exp\left(-\frac{1}{T} \sum_{\langle i,j \rangle} \delta(l_i, l_j)[1 - 2p_{ij}] \right) \quad (4) \end{aligned}$$

where $Z_1 = Z/\exp(\Upsilon)$ is a constant partition function (with Υ a factor which depends only on the data), namely

$$Z_1 = \sum_{s\in\Omega} \exp\left(-\frac{1}{T}\sum_{\langle i,j\rangle}\delta(l_i,l_j)[1-2p_{ij}]\right)$$

where Ω is the set of all possible (configurations for the) segmentations into regions ($S=\{l_i\}$) of size N pixels. Let us add that, since the number of classes (and thus the number of regions) of this final segmentation is not *a priori* known, there are possibly, between one and as much as regions that the number N of pixels in this image (assigning each pixel to an individual region is a possible configuration). In this setting, $(1-2p_{ij})$ can be viewed as the potential of spatially variant binary cliques (or pairwise interaction potentials) of an equivalent nonstationary MRF generative model of *correct* segmentations in the case where $\{S_k^{gt}\}$ is assumed to be a set of representative ground truth segmentations. Besides, S, the segmentation result (to be compared to $\{S_k^{gt}\}$), can be considered as a realization of this generative model with PRand, a statistical measure proportional to its negative likelihood energy. In other words, an estimate of S, in the maximum likelihood sense of this generative model, will give a resulting segmented map (i.e., a fusion result) with a high fidelity to the set of segmentations $\{S_k^{gt}\}$ to be fused.

D. Label Field Fusion Model for Image Segmentation

Let us consider that we have at our disposal, a set of L segmentations $\{S_1, S_2, \ldots, S_L\}$ associated to an image of size N to be fused (i.e., to efficiently combine) in order to obtain a final reliable and accurate segmentation result. The generative Gibbs distribution model of correct segmentations expressed in (4) gives us an interesting fusion model of segmentation maps, in the maximum PRI sense, or equivalently in the maximum likelihood (ML) sense for the underlying Gibbs model expressed in (4).

In this framework, the set of $\{p_{ij}\}$ is computed with the empirical proportion estimator [see (3)] on the data $\{S_1, S_2, \ldots, S_L\}$. Once $\{p_{ij}\}$ has been estimated, the resulting ML fusion segmentation map $\hat{S}_{\text{fusion}}$ is thus defined by maximizing the likelihood distribution $P(\{p_{ij}\}|S)$

$$\begin{aligned}\hat{S}_{\text{ML}_{\text{fusion}}} &= \arg\max_S P\left(\{p_{ij}\}|S\right)\\ &= \arg\min_S \left\{\sum_{\langle i,j\rangle}\delta(l_i,l_j)[1-2p_{ij}]\right\}\\ &= \arg\min_S U_{\text{L}}\left(\{p_{ij}\}, S=\{l_i\}\right) \end{aligned} \quad (5)$$

where U_{L} is the likelihood energy term of our generative fusion model which has to be minimized in order to find $\hat{S}_{\text{ML}_{\text{fusion}}}$.

Concretely, U_{L} encodes the set of constraints, in terms of pairs of pixel labels (identical or not), provided by each of the L segmentations to be fused. The minimization of U_{L} finds the resulting segmentation which also optimizes the PRI criterion.

E. Bayesian Fusion Model for Image Segmentation

As previously described in Section II-B, the image segmentation problem is an ill-posed problem exhibiting multiple solutions for different possible values of the number of classes which is not *a priori* known. To render this problem well-posed with a unique solution, some constraints on the segmentation process are necessary, favoring over segmentation or, on the contrary, merging regions. From the probabilistic viewpoint, these regularization constraints can be expressed by a prior distribution of the unknown segmentation $S=\{l_i\}$ treated as a realization of a random field, for example, within a MRF framework [2], [27] or analytically, encoded *via* a local or global [13], [28] prior energy term added to the likelihood term.

In this framework, we consider an energy function that sets a particular global constraint on the fusion process. This term restricts the number of regions (and indirectly, also penalizes small regions) in the resulting segmentation map. So we consider the energy function

$$U_{\text{P}}(S) = |\mathcal{R}(S)| \cdot H\left(|\mathcal{R}(S)| - \tau\right) \quad (6)$$

where $|\mathcal{R}(S)|$ designates the number of regions (set of connected pixels belonging to the same class) in the segmented image S, $H(.)$ is the Heaviside (or unit step) function, and τ an internal parameter of our fusion model which physically represents the number of classes above which this prior constraint, limiting the number of regions, is taken into account. From the probabilistic viewpoint, this regularization constraint corresponds to a simple shifted (from τ) exponential distribution decreasing with the number of regions displayed by the final segmentation.

In this framework, a regularized solution corresponds to the *maximum a posteriori* (MAP) solution of our fusion model, i.e., the solution $\hat{S}_{\text{MAP}_{\text{fusion}}}$ that maximizes the posterior distribution $P(S|\{p_{ij}\}) \propto P(S)P(\{p_{ij}\}|S)$, and thus

$$\begin{aligned}\hat{S}_{\text{MAP}_{\text{fusion}}} &= \arg\max_S P\left(S=\{l_i\}|\{p_{ij}\}\right)\\ &= \arg\min_S \left\{U_{\text{L}}|\left(\{p_{ij}\}, S=\{l_i\}\right) + \beta U_{\text{P}}(S)\right\}\\ &= \arg\min_S \quad \mathbf{U}\left(S=\{l_i\}, \{p_{ij}\}, \beta, \tau\right) \end{aligned} \quad (7)$$

with β is the regularization parameter controlling the contribution of the two terms; U_{L} expressing fidelity to the set of segmentations to be fused and U_{p} encoding our prior knowledge or beliefs concerning the types of acceptable final segmentations as estimates (segmentation with a number of limited regions). In this way, the resulting criteria used in this resulting fusion model can be viewed as a penalized maximum rand estimator.

III. Coarse-to-Fine Optimization Strategy

A. Multiresolution Minimization Strategy

Our fusion procedure of several label fields emerges as an optimization problem of a complex non-convex cost function U with several local extrema over the label parameter space. In order to find a particular configuration of $\hat{S}_{\text{fusion}}$, that efficiently minimizes this complex energy function, we can use a global optimization procedure such as a simulated annealing algorithm [27] whose advantages are twofold. First, it has the capability of avoiding local *minima*, and second, it does not require a good initial guess in order to estimate the $\hat{S}_{\text{fusion}}$ solution.

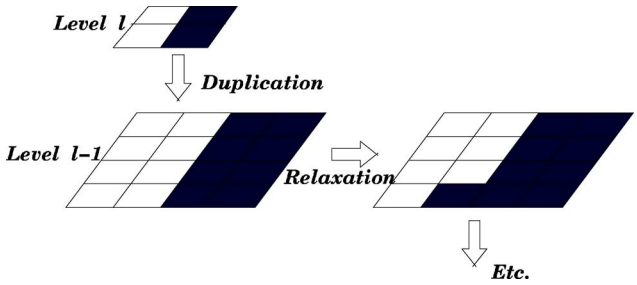

Fig. 1. Duplication and "coarse-to-fine" minimization strategy.

An alternative approach to this stochastic and computationally expensive procedure is the iterative conditional modes (ICM) introduced by Besag [2]. This method is deterministic and simple, but has the disadvantage of requiring a proper initialization of the segmentation map close to the optimal solution. Otherwise it will converge towards a bad local minima associated with our complex energy function $U()$. In order to solve this problem, we could take, as initialization (first iteration), the segmentation map $\hat{S}^{[0]}_{\text{fusion}}$ such as

$$\hat{S}^{[0]}_{\text{fusion}} = \arg \min_{S \in \{S_1, S_2, \ldots, S_L\}} \mathbf{U}(S = \{l_i\}, \{p_{ij}\}, \beta, \tau) \quad (8)$$

i.e., in choosing for the first iteration of the ICM procedure amongst the L segmentation to be fused, the one closest to the optimal solution of the Gibbs energy function of our fusion model [see (5)].

A more robust optimization method consists of a multiresolution approach combined with the classical ICM optimization procedure. In this strategy, rather than considering the minimization problem on the full and original configuration space, the original inverse problem is decomposed in a sequence of approximated optimization problems of reduced complexity. This drastically reduces computational effort and provides an accelerated convergence toward improved estimate. Experimentally, estimation results are nearly comparable to those obtained by stochastic optimization procedures as noticed, for example, in [10] and [29].

To this end, a multiresolution pyramid of segmentation maps for each S_k $k = 0, \ldots, L$ is preliminarily derived, in order to estimate a set of $\{p_{ij}\}$ at different resolution levels, and a set of *similar* spatial models is considered for each resolution level of the pyramidal data structure. At the upper level of the pyramidal structure (lower resolution level), the ICM optimization procedure is initialized with the segmentation map given by the procedure defined in (8). It may also be initialized by a random solution and, starting from this initial segmentation, it iterates until convergence. After convergence, the result obtained at this resolution level is interpolated (see Fig. 1) and then used as initialization for the next finer level and so on, until the full resolution level.

B. Optimization of the Full Energy Function

Experiments have shown that the full energy function of our model, (with the region based-global regularization constraint) is complex for some images. Consequently it is preferable to perform the minimization in two steps. In a first step, the minimization is performed without considering the global constraint (considering only U_{L}), with the previously mentioned multiresolution minimization strategy and the ICM optimization procedure until its convergence at full resolution level. At this finest resolution level, the minimization is then refined in a second step by identifying each region of the resulting segmentation map. This creates a region adjacency graph (a RAG is an undirected graph where the nodes represent connected regions of the image domain) and performs a region merging procedure by simply applying the ICM relaxation scheme on each region (i.e., by merging the couple of adjacent regions leading to a reduction of the cost function of the full model [see (7)] until convergence). In the second step, minimization can also be performed according to the full model ($\beta \neq 0$).

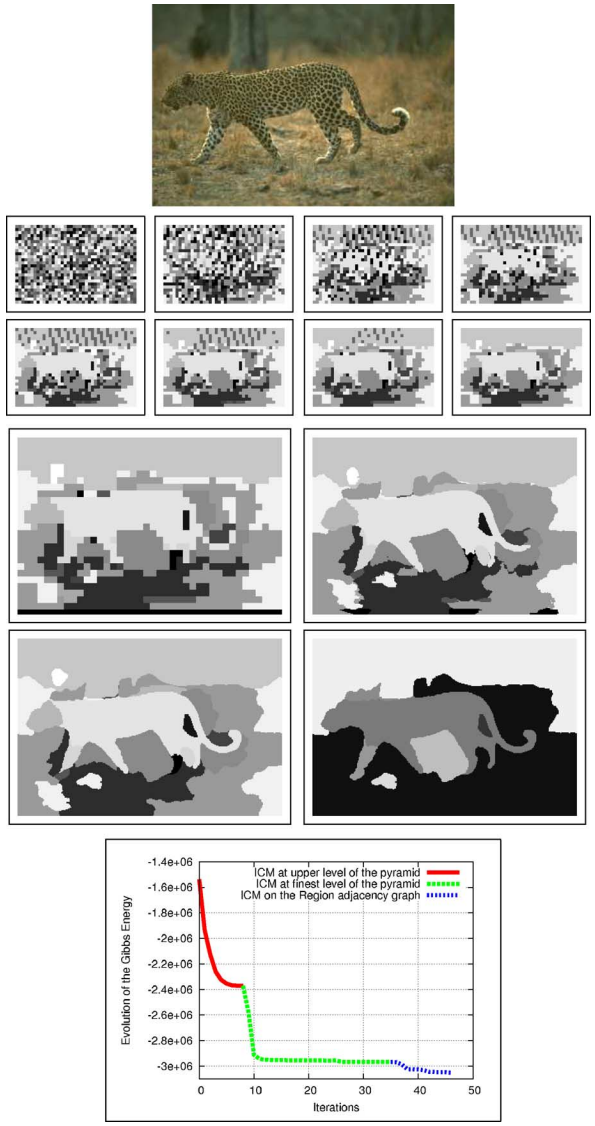

Fig. 2. From top to bottom and left to right; A natural image from the Berkeley database (no. 134052) and the formation of its region process (algorithm $\text{PRIF}_{[K_1=16|K_2=10|K_3=2]}$) at the ($l = 3$) upper level of the pyramidal structure at iteration [0–6], 8 (the last iteration) of the ICM optimization algorithm. Duplication and result of the ICM relaxation scheme at the finest level of the pyramid at iteration 0, 1, 18 (last iteration) and segmentation result (region level) after the merging of regions and the taking into account of the prior. Bottom: evolution of the Gibbs energy for the different steps of the multiresolution scheme.

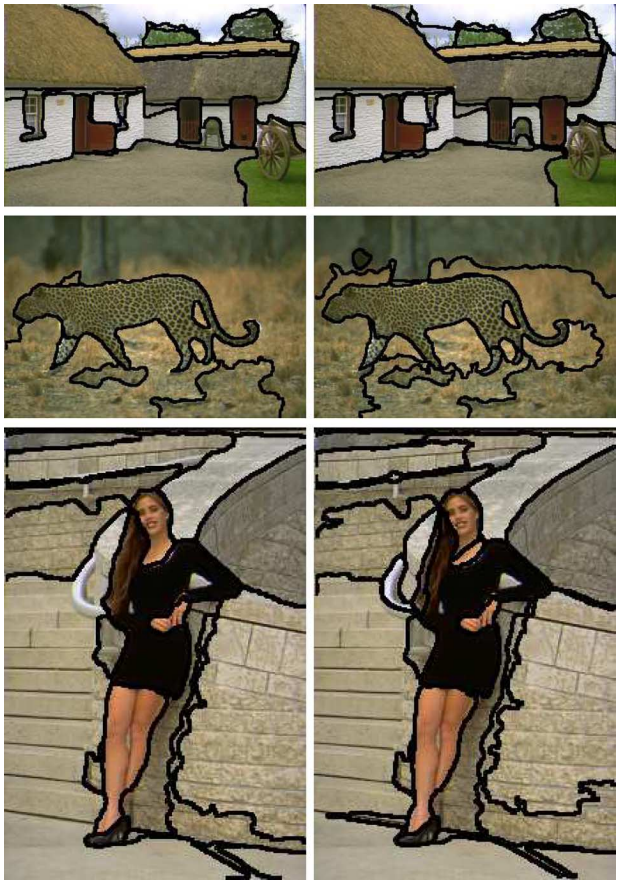

Fig. 3. Comparison of two segmentation results of our multiresolution fusion procedure (algorithm $\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2]}$) using respectively: left] a subsampled and fixed number of connections (85) regularly spaced and located within a square *search window* of size $N_s = 30$ pixels. right] a fully connected graph computed on a *search window* two times larger (and requiring a computational load increased by 100).

Fig. 4. Segmentation (image no. 385028 from Berkeley database). From top to bottom and left to right; segmentation map respectively obtained by 1] our multiresolution optimization procedure: $U = \mathbf{-3402965}$ (algorithm $\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2]}$), 2] $\mathrm{SA}_{[K_{\max}=100]}$: $U = -3206127$, 3] $\mathrm{SA}_{[K_{\max}=500]}$: $U = -3312794$, 4] $\mathrm{SA}_{[K_{\max}=1000]}$: $U = -3395572$, 5] $\mathrm{SA}_{[K_{\max}=5000]}$: $U = -3402162$.

C. Algorithm

In order to decrease the computational load of our multiresolution fusion procedure, we only use two levels of resolution in our pyramidal structure (see Fig. 2): the full resolution and an image eight times smaller (i.e., at the third upper level of classical data pyramidal structure). We do not consider a *complete* graph: we consider that each node $\mathbf{x}_i$ (or pixel) is connected with its four nearest neighbors and a fixed number of connections (85 in our application), regularly spaced between all other pixels located within a square *search window* of fixed size $N_s = 30$ pixels centered around $\mathbf{x}_i$. Fig. 3 shows comparison of segmentation results with a fully connected graph computed on a *search window* two times larger.

We decided to initialize the lower (or third upper) level of the pyramid with a sequence of 20 different random segmentations with K_1 classes. The full resolution level is then initialized with the duplication (see Fig. 1) of the best segmentation result (i.e., the one associated to the lowest Gibbs energy U_L) obtained after convergence of the ICM at this lower resolution level (see Fig. 2). We provide details of our optimization strategy in Algorithm 1.

Algo I. Multiresolution minimization procedure (see also Fig. 2).

Two-Step Multiresolution Minimization

$\{S_k\}$	Set of L segmentations to be fused
$\{p_{ij}\}^{[l]}$	Pairwise probabilities for each pixel pair computed over $\{S_k\}$ at resolution level l

1. Initialization Step

- $\{\mathcal{P}_k\} \leftarrow$ Build L multiresolution Pyramids from $\{S_k\}$
- $\{p_{ij}\}^{[3]} \leftarrow$ Compute the pairwise probabilities from $\{\mathcal{P}_k\}$ at resolution level 3
- $\{p_{ij}\}^{[0]} \leftarrow$ Compute the pairwise probabilities from $\{S_k\}$ at full resolution ($\{p_{ij}\}^{[0]} = \{p_{ij}\}$)

PIXEL LEVEL

Initialization: *Random initialization of the upper level of the pyramidal structure with K_1 classes*

- ICM optimization on $U_{\mathrm{L}}(\{p_{ij}\}^{[l=3]})$
- Duplication (cf. Fig 1) to the full resolution
- ICM optimization on $U_{\mathrm{L}}(\{p_{ij}\})$

REGION LEVEL

for *each region at the finest level* **do**

- ICM optimization on $\mathbf{U}(\{p_{ij}\}, \beta, \tau) = U_{\mathrm{L}}(\{p_{ij}\}) + \beta U_{\mathrm{P}}(\tau)$

D. Comparison With a Monoresolution Stochastic Relaxation

In order to test the efficiency of our two-step multiresolution relaxation (MR) strategy, we have compared it to a standard monoresolution stochastic relaxation algorithm, i.e., a so-called simulated annealing (SA) algorithm based on the Gibbs sampler [27]. In order to restrict the number of iterations to be finite, we have implemented a geometric temperature cooling schedule [30] of the form $T_k = T_0 \cdot (T_f/T_0)^{k/K_{\max}}$, where T_0 is the starting temperature, T_f is the final temperature, and $K_{\max}$ is the maximal number of iterations. In this stochastic procedure, the choice of the initial temperature T_0 is crucial. The temperature must be sufficiently high in the first stages of simulated

TABLE I
GIBBS ENERGY VALUES OBTAINED BY OUR MULTIRESOLUTION RELAXATION (MR) METHOD AND BY THE SIMULATED ANNEALING (SA) PROCEDURE ALONG WITH $K_{\max} \in \{100, 500, 1000, 5000\}$

	Final Energy U			
ALGORITHMS	Image 388016	Image 385039	Image 385028	Image 134052
MR	-2948922	$\mathbf{-3483070}$	$\mathbf{-3402965}$	$\mathbf{-2867114}$
$\mathrm{SA}_{[K_{\max}=0.1K]}$	-2813774	-3367534	-3206127	-2738007
$\mathrm{SA}_{[K_{\max}=0.5K]}$	-2943968	-3395013	-3312794	-2837538
$\mathrm{SA}_{[K_{\max}=1K]}$	-2947328	-3375435	-3395572	-2831235
$\mathrm{SA}_{[K_{\max}=5K]}$	$\mathbf{-2962743}$	-3462671	-3402162	-2846963

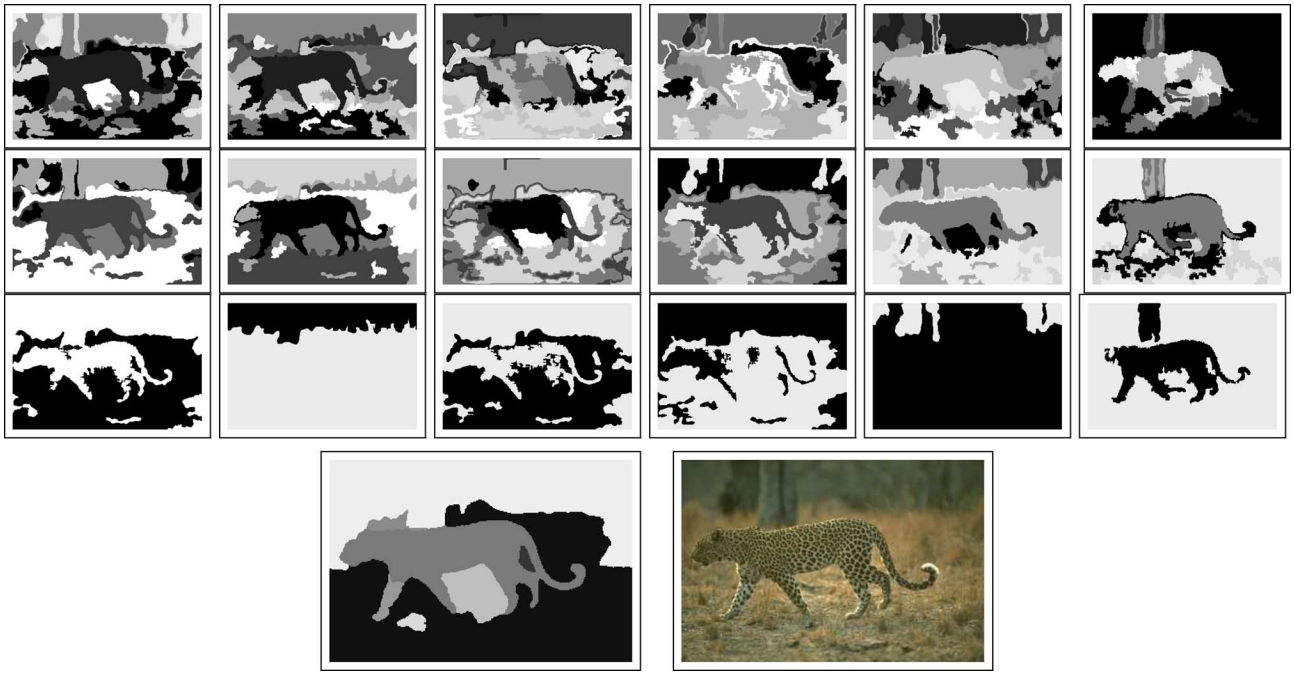

Fig. 5. Examples of fusion results (algorithm $\mathrm{PRIF}_{[K_1=16|K_2=10|K_3=2]}$). Three first rows; K-mean clustering results for the segmentation model described in Section IV (expressed in the RGB, HSV, YIQ, XYZ, LAB, and LUV color spaces) into $K_1 = 12$, $K_2 = 6$ and $K_3 = 2$ classes (with $q_b = 5$) for respectively the first, second and third row. Input natural image from the Berkeley image database and final segmentation map resulting of the fusion by the $\mathrm{PRIF}_{[K_1=16|K_2=10|K_3=2]}$.

annealing in order to widely explore the solution space. Experimentally, we have noticed that a reliable initial temperature is given by $T_0 = 40$. This ensures a good exploration in all tested cases, i.e., that at least 50% of sites change their class during the first iteration. We have also noticed that a good final temperature, ensuring that less than 0.05% of sites change their class between two complete image sweeps, is $T_{\mathrm{f}} = 0.25$.

Table I summarizes (for the three last images of the Berkeley database plus the image presented in Fig. 2) the Gibbs energy values obtained by our multiresolution minimization method and by the simulated annealing procedure along with different values of $K_{\max} \in \{100, 500, 1000, 5000\}$. Fig. 4 compares some segmentation results obtained by these two optimization methods. We can observe that the proposed multiresolution approach gives results, in terms of Gibbs energy values and segmentation results, close to a simulated annealing with $K_{\max}$ =1000–5000 (and often more) iterations, i.e., corresponding to a gain of almost two orders of magnitude in terms of iterations or computational cost comparatively to our multiresolution approach (requiring less than 50 iterations: see Fig. 2).

IV. SEGMENTATIONS TO BE FUSED

The initial segmentation maps which will be fused by our fusion framework are simply given, in our application, by a K-means [31] clustering technique, applied on an input image expressed by different color spaces and different number of classes. As simple cues (i.e., as input multidimensional feature descriptor), we used the set of values of the re-quantized color histogram, with equidistant binning, estimated around the pixel to be classified. In our application, this local histogram is equally requantized, for each of the three color channels, in a $N_b = q_b^3$ bin descriptor, computed on an overlapping squared fixed-size ($N_w = 7$) neighborhood centered around the pixel to be classified (see Algorithm 2 and [14] for more details).

Algo II. Bin descriptor estimation for each pixel.

Estimation of the bin descriptor for each pixel

$\mathcal{N}_{\mathbf{x}}$	Set of pixel locations $\mathbf{x}$ within the $N_w \times N_w$ neighborhood region centered at $\mathbf{x}$
$h[\,]$	Bin descriptor: array of N_b floats $(h[0], h[1], \cdots, h[N_b-1])$
$\lfloor . \rfloor$	Integer part of.

for *each pixel* $\mathbf{x} \in \mathcal{N}_{\mathbf{x}}$ *with color value* $R_{\mathbf{x}}, G_{\mathbf{x}}, B_{\mathbf{x}}$ **do**

- $k \leftarrow q^2 \cdot \lfloor q \cdot R_{\mathbf{x}}/256 \rfloor + q \cdot \lfloor q \cdot G_{\mathbf{x}}/256 \rfloor + \lfloor q \cdot B_{\mathbf{x}}/256 \rfloor$
- $h[k] \leftarrow h[k] + 1/N_w^2$

In this simpler model, a *texton* (i.e., the repetitive character or element of a textured image, also called a texture primitive) is herein characterized by a mixture of colors, or more precisely, by the values of the requantized local color histogram. This model is simple to compute, allows significant data reduction while being robust to noise and local image transformations and has already demonstrated efficiency for tracking applications [32].

Finally, these (q_b^3-bin) descriptors are grouped into different clusters, corresponding to each class of the image, by the classical K-means algorithm [31] with the L_1 norm (also called Manhattan distance). A final merging step is added to each segmentation map that simply consists of fusing each small region (i.e., regions whose size is below 200 pixels) with the region sharing its longest boundary.

The execution time of each segmentation map provided by this simple segmentation model remains fast and, depending on K and q, requires between 0.25 and 2 s for a non-optimized C++ code.

This simple segmentation strategy of the input image into K classes is repeated for different color spaces (which can be viewed as different image channels provided by various sensors or captors), and for different values of K (the number of classes) and finally for different values of q_b (i.e., different values of the number of bins of the requantized color histogram).

Each color space has an interesting property, which can efficiently be taken into account in order to render final fusion procedure more reliable [1], [33].

For example, RGB is an additive color system based on trichromatic theory and this color system is nonlinear with visual perception. This color space seems to be the optimal one for tracking applications [34]. The HSV is interesting in order to decouple chromatic information from shading effect [32]. Due to its excellent decorrelation property, discriminating power and transmission efficiency, the YIQ system is also intended to take advantage of human color-response characteristics. These properties explain why this color system is also used by the NTSC color TV system. XYZ has the advantage of being more psycho-visually linear although it is nonlinear in terms of linear component color mixing. The LAB color system approximates human vision, and its L component closely matches human perception of lightness. The LUV components provide an Euclidean color space yielding a perceptually uniform spacing of color approximating a Riemannian space [35] (the euclidean distance between two points in this space indicates more or less the perceptual difference between them). The $I_1I_2I_3$ (or Otha color's space) is the approximation of the Karhunen Loeve transform of the RGB space. It creates color channels which are orthogonal and completely decorrelated. Consequently, this color system has the property to be optimal in terms of energy compaction and has shown invariance with respect to highlights and illumination intensity. The $H_1H_2H_3$ [36] is the only color space which is only based on channel differences. It is only invariant with respect to highlights, a useful feature in image understanding when observing shiny surfaces [33]. Finally, TSL color space is very effective for skin segmentation and detection [37] when using a Gaussian model (even if illumination conditions vary).

The different values of the number of classes take into account the inherent variability of the possible interpretations between human observers of an image, i.e., the multiple acceptable levels of details with which a human choose to segment a natural image. The different values of the number of bins can be viewed as the output result of different image channels giving the same information at different resolution levels.

Of course, these initial segmentations to be fused can be provided by different segmentation models, or different segmentation results provided by different seeds of the same stochastic segmentation model, different results of the same deterministic segmenter with different initializations or different images provided by different channels or sensors or finally provided by the output result of a K-mean clustering with several different metrics (which is *somewhat* similar to considering different color spaces).

V. Experimental Results

A. Set Up

In all the experiments, we have considered our fusion methods on initial segmentations obtained with the simple model segmentation presented in Section IV with the following parameters: the size of the squared window, used to compute the local histogram for the initial segmentations is set to $N_w = 7 \times 7$.

Experiments have shown that the proposed fusion model is more efficient if the initial segmentations to be fused are diversified. In order to take this empirical observation into account, we thus consider the following for a total of 60 segmentations to be fused.

1) Ten $(N_s = 10)$ different color spaces, namely RGB, HSV, YIQ, XYZ, LAB, LUV, $\mathrm{I_1I_2I_3}$, $\mathrm{H_1H_2H_3}$, $\mathrm{YC_bC_r}$, TSL [1], [38], [39].
2) Three different values of the number of classes, respectively K_1, K_2 and K_3.
3) Two different values of the number of bins for each local re-quantized color histogram, namely $N_b = 5^3$ and $N_b = 4^3$.

B. Performance Measures and Comparison With State-of-the-Art Methods

We have replicated the scenario used in the evaluation of state-of-the-art segmentation methods described in [40]–[42][1] and in [43]. In these experiments, we have to test our segmentation algorithm on the Berkeley segmentation database [26] consisting of 300 color images of size 481×321 divided into a training set of 200 images, and a test set of 100 images. For each color image, a set of benchmark segmentation results, provided by human observers (between 4 and 7), is available and will be used to quantify the reliability of the proposed segmentation algorithm. In order to ensure the integrity of the evaluation, the internal parameters of the algorithm (K_1, K_2, K_3, β, and τ) are tuned on the train image set by doing a local discrete grid search routine, with a fixed step-size, on the parameter space and in the feasible ranges of parameter values (namely $K_1, K_2, K_3 \in [2-20]$ [step-size: 2], $\beta \in [0-4000]$ [step-size: 1000], $\tau \in [5-20]$ [step-size: 5]). Those presented in Tables II and III are only a few examples. The algorithm is then bench-marked by using the optimal training parameters on the independent test set. We have compared our segmentation algorithm (called $\mathrm{PRIF}_{[K_1|K_2|K_3][\beta,\tau]}$ for PRIF, K_1, K_2, K_3, β, and τ being its five internal parameters) against five unsupervised algorithms which are publicly available. For each of these algorithms, the internal parameters are set to optimal values (see [40]) and/or correspond to the internal values suggested by the authors. These algorithms are namely the mean-shift [3] (with $h_s = 13$, $h_r = 19$), Ncuts [5] (with a number of segments $K = 20$, agreeing with the average number of regions found in the segmentation maps given by the human observers [41]), and FH [6] (with a smoothing parameter $\sigma = 0.5$, a threshold value $k = 500$ and a minimal region size equals to 200 pixels) and the CTM algorithm proposed in [40], [42] (with $\eta = 0.1$ and $\eta = 0.2$) and the FCR [14] for two sets of internal parameters and finally the CTex [44] and the JSEG [45]. As in [40], [41], all color images are normalized to have the longest side equals to 320 pixels. The segmentation results are then supersampled in order to obtain segmentation images with the original resolution (481×321) before the estimation of the performance metrics.

The comparison is based on the following performance measures, namely the PRI measure (higher probability is better and a score equal to $\mathrm{PRI} = 0.80$ means that on average 80% of pairs of pixel labels are correctly classified in the segmentation results) which seems to be highly correlated with human hand-segmentations [40] along with the F-measure proposed by Martin *et al.* [43], [46], [47]. This F-measure, deduced from the Precision/Recall values and characterizing the agreement between region boundaries of two segmentations, is now widely used in the computer vision and edge detection community. Qualitatively, the precision measure (P) is defined as the fraction of detections that are true boundaries; this measure is low when there is significant over-segmentation, or when a large number of boundary pixels have poor localization. The Recall (R) measure gives the fraction of true boundaries detected; a low recall value is typically the result of under-segmentation and indicates failure to capture salient image structure. Thus, Precision quantifies the amount of noise in the output of a detector, while Recall quantifies the amount of ground-truth detected. The performance of a boundary detector providing a binary output is represented by a point in the Precision-Recall plane. If the output is a soft boundary image, a parametric Precision-Recall curve expresses the compromise between absence of noise and fidelity to ground truth as the main parameter of the boundary detector varies. Precision and Recall can be combined in a single quality statistic measure, the F-measure ($F = 2PR/(P+R)$), defined as their harmonic mean and as a measure of performance combining both precision and recall. The maximal F-measure on a Precision-Recall curve is used as a summary statistic for the performance of the detector on a set of images.

These two performance measures have to be considered together and a good segmentation algorithm will give good scores for these two complementary performance measures.

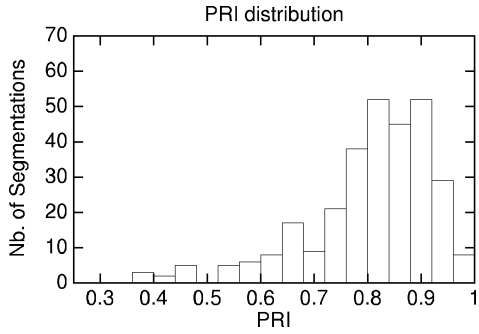

Fig. 6. Distribution of the PRI performance measure over the 300 images of the Berkeley database (for $\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2]}$).

C. Results

Tables II and III show the obtained results for different values of K_1, K_2, K_3, β, and τ. The first five tested algorithms are with $\beta = 0.0$ (without prior term).

In terms of PRI measure, we observe that the discussed fusion strategy gives competitive results (a score equals to $\mathrm{PRI} = 0.80$ means that on average 80% of pairs of pixel labels are correctly classified in the segmentation results) with a relative low variance over the set of images of the Berkeley image database. Fig. 6 shows the distribution of the PRI measure over the 300 images of the Berkeley image database for the algorithm $\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2]}$.

In terms of F-measure, we can observe that the obtained segmentations give a lower F-measure compared to the best existing boundary detection algorithms whose the highest score to date [48] is $F = 0.70@(0.71, 0.68)$[2] [47]. Let us note that our algorithm also gives a set of closed curves (by giving a "hard" boundary representation) compared to most of the boundary detectors benchmarked in [47]. For this reason, our algorithm have to be compared with the best boundary detection algorithm ensuring a segmentation map (i.e., a set of closed curves) such as the one proposed in [8] whose F-measure is $F = 0.67@(0.69, 0.66)$.

[1]We have used the MATLAB source code, proposed by Yang in order to estimate the PRI performance measure presented in the following Section. This code is kindly available online at http://www.eecs.berkeley.edu/~yang/software/lossy_segmentation/.

[2]The notation F@(recall,precision) represents the value of the highest F performance measure (of a binary classifier) existing on its precision-recall (or ROC) curve at coordinates (@), the measure of its recall performance (on x-axis) and its precision performance (on y-axis).

TABLE II
AVERAGE PERFORMANCE, IN TERMS OF PRI MEASURE, OF OUR ALGORITHM FOR SEVERAL VALUES OF ITS INTERNAL PARAMETERS ON THE BERKELEY IMAGE DATABASE [26]. THE FIRST VALUE IS THE PERFORMANCE METRIC ON THE ENTIRE DATABASE AND VALUES BETWEEN SQUARE BRACKETS CORRESPOND TO PERFORMANCES ON THE TRAIN AND TEST IMAGE SETS

ALGORITHMS	PRI [16]
HUMANS (in [40])	0.8754
PRIF$_{[K_1=18\|K_2=10\|K_3=2]}$	**0.8006** [train ▷ **0.8062** test ▷ **0.7894**]
PRIF$_{[K_1=16\|K_2=10\|K_3=2]}$	0.7986 [train ▷ 0.8057 test ▷ 0.7846]
PRIF$_{[K_1=13\|K_2=8\|K_3=2]}$	0.7982 [train ▷ 0.8053 test ▷ 0.7838]
PRIF$_{[K_1=18\|K_2=10\|K_3=4]}$	0.7899 [train ▷ 0.7952 test ▷ 0.7792]
PRIF$_{[K_1=18\|K_2=10\|K_3=6]}$	0.7848 [train ▷ 0.7909 test ▷ 0.7727]
PRIF$_{[K_1=18\|K_2=10\|K_3=8]}$	0.7809 [train ▷ 0.7867 test ▷ 0.7692]
PRIF$_{[K_1=18\|K_2=10\|K_3=2][\beta=3000\|\tau=10]}$	0.7940 [train ▷ 0.8015 test ▷ 0.7791]
CTex [44]	0.8000
FCR$_{[K_1=13\|K_2=6\|\kappa=0.135]}$ [14]	0.7882
FCR$_{[K_1=13\|K_2=13\|\kappa=0.145]}$ [14]	0.7849
FH [6] (in [40])	0.7841
JSEG [45] (in [44])	0.7700
CTM$_{\eta=0.15}$ [40][44]	0.7627
CTM$_{\eta=0.20}$ [40][44]	0.7617
CTM$_{\eta=0.10}$ [40][44]	0.7561
Mean-Shift [3] (in [40])	0.7550
NCuts [5] (in [40])	0.7229

TABLE III
AVERAGE PERFORMANCE, IN TERMS OF F MEASURE [43], OF OUR ALGORITHM (HARD AND SOFT BOUNDARY REPRESENTATION) FOR SEVERAL VALUES OF ITS INTERNAL PARAMETERS ON THE BERKELEY IMAGE DATABASE [26] (ON THE TRAIN AND TEST IMAGE SETS)

ALGORITHMS		F Measure [43]
HUMANS		0.79
Hard	PRIF$_{[K_1=18\|K_2=10\|K_3=2]}$	train ▷ 0.61 @(0.56 , 0.68) test ▷ 0.59 @(0.54 , 0.65)
	PRIF$_{[K_1=16\|K_2=10\|K_3=2]}$	train ▷ 0.61 @(0.56 , 0.68) test ▷ 0.59 @(0.54 , 0.65)
	PRIF$_{[K_1=13\|K_2=8\|K_3=2]}$	train ▷ 0.61 @(0.55 , 0.68) test ▷ 0.59 @(0.54 , 0.65)
	PRIF$_{[K_1=18\|K_2=10\|K_3=4]}$	train ▷ **0.62 @(0.64 , 0.61)** test ▷ **0.60 @(0.62 , 0.58)**
	PRIF$_{[K_1=18\|K_2=10\|K_3=6]}$	train ▷ 0.62 @(0.67 , 0.58) test ▷ 0.60 @(0.66 , 0.55)
	PRIF$_{[K_1=18\|K_2=10\|K_3=8]}$	train ▷ 0.61 @(0.69 , 0.55) test ▷ 0.60 @(0.68 , 0.53)
	PRIF$_{[K_1=18\|K_2=10\|K_3=2][\beta=3000\|\tau=10]}$	train ▷ 0.60 @(0.54 , 0.69) test ▷ 0.58 @(0.52 , 0.67)
Soft	PRIF$_{[K_1=18\|K_2=10\|K_3=2]}$	train ▷ 0.62 @(0.54 , 0.72) test ▷ 0.60 @(0.53 , 0.69)
	PRIF$_{[K_1=16\|K_2=10\|K_3=2]}$	train ▷ 0.62 @(0.54 , 0.72) test ▷ 0.60 @(0.53 , 0.68)
	PRIF$_{[K_1=13\|K_2=8\|K_3=2]}$	train ▷ 0.62 @(0.53 , 0.73) test ▷ 0.59 @(0.53 , 0.68)
	PRIF$_{[K_1=18\|K_2=10\|K_3=4]}$	train ▷ 0.64 @(0.61 , 0.68) test ▷ 0.62 @(0.59 , 0.65)
	PRIF$_{[K_1=18\|K_2=10\|K_3=6]}$	train ▷ **0.64 @(0.63 , 0.66)** test ▷ **0.63 @(0.62 , 0.63)**
	PRIF$_{[K_1=18\|K_2=10\|K_3=8]}$	train ▷ 0.64 @(0.64 , 0.64) test ▷ 0.63 @(0.62 , 0.64)
	PRIF$_{[K_1=18\|K_2=10\|K_3=2][\beta=3000\|\tau=10]}$	train ▷ 0.61 @(0.53 , 0.73) test ▷ 0.59 @(0.51 , 0.70)

Let us recall that a "soft" output is more appropriate for this measure, since the benchmark algorithm (for computing the F measure) will also find the optimal threshold without requiring a set of closed contours. Comparison between our segmentation results and the ground-truths provided by the Berkeley database shows that our method tends to give an over-segmentation, thus inducing some noise (false contours) in the final segmentations. In order to validate this assumption and to objectively compare our segmentation method with the state-of-the-art boundary detectors (generally giving a "soft" output whose thresholding operation does not ensure a set of closed contours) we have tested the use of our final "hard" boundary representation converted to a soft version, simply replacing each detected boundary pixel by the numerical gradient magnitude of its N_b-bin descriptor (i.e., the requantized local color histogram computed in a small squared-size window on the initial color image, see Section IV and Algorithm 2). Numerically speaking, by computing for each pixel

TABLE IV
MEAN NUMBER OF REGIONS IN THE K-MEANS SEGMENTATIONS TO BE FUSED AND IN THE FINAL SEGMENTATION FOR THE FUSION MODEL WITHOUT PRIOR TERM

ALGORITHMS	Mean number of regions	
	K-means	Final
$\mathrm{PRIF}_{[K_1=18\|K_2=10\|K_3=2]}$	33.62	16.95
$\mathrm{PRIF}_{[K_1=16\|K_2=10\|K_3=2]}$	33.12	16.94
$\mathrm{PRIF}_{[K_1=13\|K_2=8\|K_3=2]}$	30.39	16.05
$\mathrm{PRIF}_{[K_1=18\|K_2=10\|K_3=4]}$	37.80	27.58
$\mathrm{PRIF}_{[K_1=18\|K_2=10\|K_3=6]}$	41.04	34.10

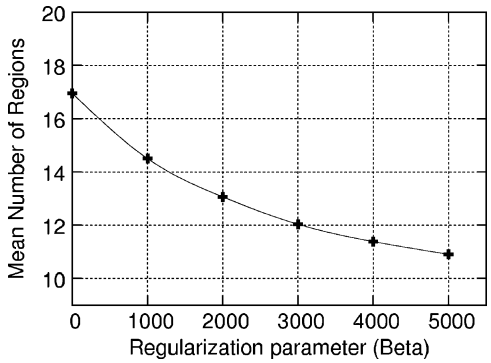

Fig. 7. Mean number of regions of $\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2][\beta|\tau=10]}$, our penalized fusion model, as a function of the regulation parameter β. We recall that the number of regions in the K-means segmentations to be fused is 33.6 in all these fusion results.

Fig. 8. First row and from left to right; a natural image from the Berkeley database (no. 229036) and the resulting fusion map using the set of 7 input hand-labeled ground-truth segmentations of the Berkeley database [26] with their number of regions and their Gibbs energy (the two first outliers related to the two first maximal energy are indicated in bold).

(detected as boundary point), located at row i and column j, the following simple distance

$$\sum_{\mathcal{C}} \left(\mathcal{D}\left(h_{i-(N_w/2),j}, h_{i+(N_w/2),j}\right) + \mathcal{D}\left(h_{i,j-(N_w/2)}, h_{i,j+(N_w/2)}\right)\right) \quad (9)$$

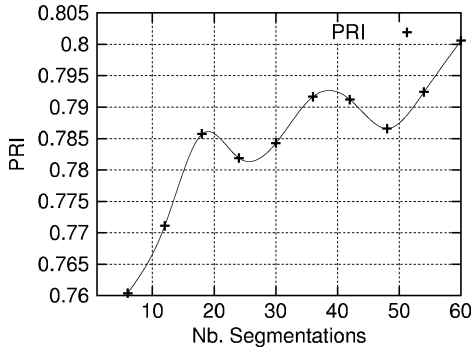

Fig. 9. Evolution of the PRI measures (averaged on the entire Berkeley database) as a function of the number of segmentations (N_s) to be fused for the $\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2]}$ algorithm. More precisely for $N_s = 6, 12, 18, 24, \ldots, 60$ segmentations (by considering only RGB space. only (RGB, HSV) spaces, only (RGB, HSV, YIQ) spaces, etc.).

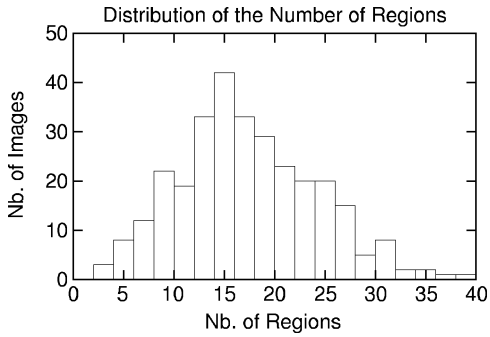

Fig. 10. Distribution of the number of regions obtained by our algorithm ($\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2]}$) over the 300 images of the Berkeley database.

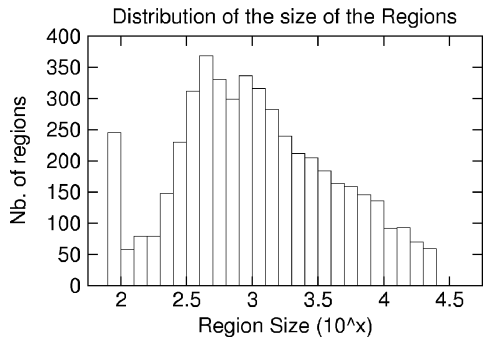

Fig. 11. Distribution of the size of the regions (log scale in x) obtained by the $\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2]}$ algorithm over the 300 images of the Berkeley database.

where the first summation is done on the ten used different color spaces, $\mathcal{D}(h_{i-d,j}, h_{i+d,j})$ is the Manhattan distance (L_1 norm) between vectors (or bin descriptors) $h_{i-d,j}$ and $h_{i+d,j}$ computed on a squared N_w-size window centered respectively at location $(i-d, j)$ and $(i+d, j)$ (see Algorithm 2). Table III shows the results which now score best with $F = 0.63@(0.62, 0.63)$.

Figs. 12 and 13 display some examples of segmentations obtained by our algorithm. The results for the entire database are available online at http://www.iro.umontreal.ca/~mignotte/ResearchMaterial/. Statistics on the mean number of regions in the K-means segmentations to be fused and in the final segmentation are summarized in Table IV for the fusion model without

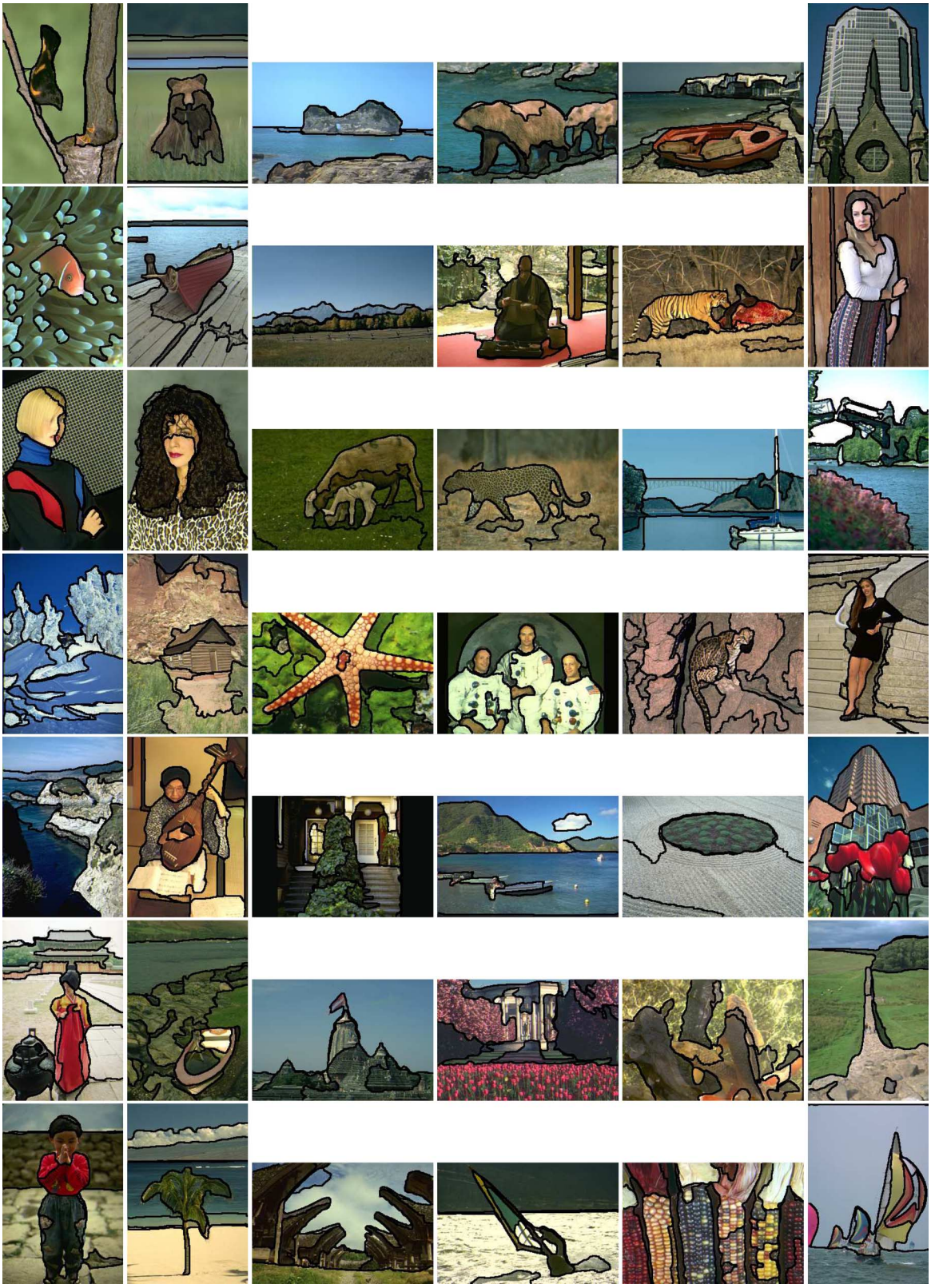

Fig. 12. Example of segmentations obtained by our algorithm $\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2]}$ on several images of the Berkeley image database (see also Tables II and III for quantitative performance measures and http://www.iro.umontreal.ca/~mignotte/ResearchMaterial/prif.html for the segmentation results on the entire database).

prior term (result averaged for the entire Berkeley database). The mean number of regions in the final segmentation as a function of the regularization parameter β, for the fusion model with prior term, i.e., $\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2][\beta|\tau=10]}$, is represented in Fig. 7. We can observe that a good PRI performance measure can be obtained without prior error term or equivalently without prior distribution constraining the number of regions of a likely segmentation as estimate. In this case, the number of

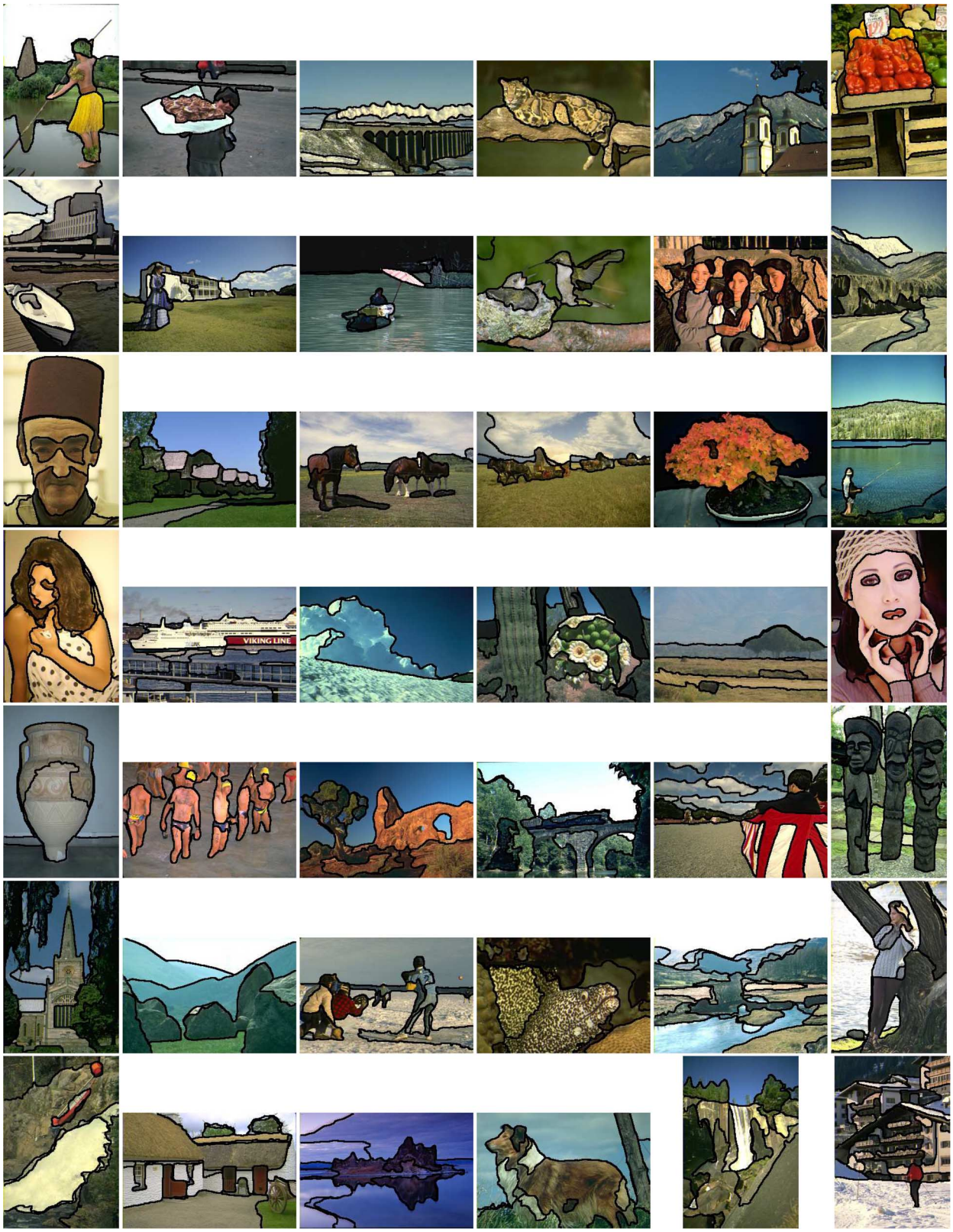

Fig. 13. Example of segmentations obtained by our algorithm $\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2]}$ on several images of the Berkeley image database (see also Tables II and III for quantitative performance measures and http://www.iro.umontreal.ca/~mignotte/ResearchMaterial/prif.html for the segmentation results on the entire database).

regions in the resulting segmentation depends on the number of regions existing in the segmentations to be fused (see Table IV). Nevertheless, our prior error term defined in (6) remains useful and interesting if we want to restrict the region number of the resulting segmentations without excessive degradation of the different performance measures (see Tables II and III). Indeed, we obtain 17 regions in average per segmentation for the $\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2]}$ and only 12 regions in average per

segmentation for the $\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2][\beta=3000|\tau=10]}$ (see Table IV and Fig. 7). This regularization term is thus not especially interesting for the application of our fusion model in the specific framework of the segmentation of natural images. However, it could be useful for other segmentation problems (e.g., hyperspectral, medical, or satellite images) for which this kind of information could be *a priori* available.

D. Discussion

We can also observe (see Fig. 9) that the PRI performance measure is better when N_s (number of segmentation to be fused) is high. The same observation can be made for the F-measure. This experiment shows the validity of our fusion procedure which is perfectible. For further improvement, we could add, to the set of segmentations to be fused, the segmentation maps obtained with $N_b = 3^3$, or the clustering results obtained by using (in the K-mean algorithm) different feature descriptors or different similarity measures between the histogram descriptors (such as the Bhattacharya, Chord, Kolmogorov, Histogram intersection, Kullkack, Shannon-Jensen distances, etc.). We could also add the results of different initializations of the K-mean clustering algorithm.

Statistics on the segmentation results by our method (e.g., distribution of the number of regions, size of the regions of the segmented Berkeley database images), for the algorithm $\mathrm{PRIF}_{[K_1=18|K_2=10|K_3=2]}$, are given in Figs. 10 and 11.

In order to further validate our fusion procedure, we experimented with the fusion of several mean-shift segmentations obtained by randomly varying h_s and h_r within respectively the intervals $[2,\ldots,15]$ and $[20,\ldots,40]$. Once again, our fusion procedure noticeably improves the performance measures in comparison of a single segmentation (see Table II). We obtain (in this case) $\mathrm{PRI} = 0.7905$.

Finally, we have also tested our fusion algorithm on the set of input hand-labeled ground-truth segmentations related to a given Berkeley image (see Fig. 8). To make this experience interesting and informative, we have chosen a natural image exhibiting a high variation existing across the possible (manually generated) segmentations. The resulting fusion map (see Fig. 8) exhibits an average number of regions, comparatively to the set of ground-truths, and can be compared to the input manually generated segmentations with the same number of regions. Let us add that the Gibbs energy of our fusion model can also be used in a second step in order to compute the fidelity of each ground-truth segmentation to the set of input ground-truths; the negative Gibbs energy being proportional to the PRand metric of each input segmentation to the resulting fusion map. This information can then be efficiently used to remove outliers or to weight the confidence, in the PRI sense, of each input segmentations (see Fig. 8).

E. Algorithm

The segmentation procedure takes, on average, less than three minutes for an AMD Athlon 64 Processor 3500+, 2.2 GHz, 4435.67 bogomips and non-optimized code running on Linux. Each one of the three steps i.e., 1] of estimations of the segmentations to be fused and 2–3] the two steps of the minimization procedure of our algorithm takes, on average, one minute for a 320×214 image. Let us add that the initial segmentations to be fused (step I.) and the proposed fusion method (step II. and III.) can be easily computed in parallel. It is straightforward for step I. of our procedure but also truth for the two last steps of our fusion model. The final Markovian energy minimization can be efficiently implemented by using the parallel abilities of a graphic processor unit (GPU) (embedded on most graphics hardware nowadays available on the market) and can be greatly accelerated as indicated in [49].

Source code (in C++ language) of our algorithm with the set of segmented images are publicly available at http://www.iro.umontreal.ca/~mignotte/ResearchMaterial/ in order to make possible eventual comparisons with future segmentation algorithms or different performance measures.

VI. Conclusion

In this paper, we have presented a new and efficient segmentation strategy based on a Markovian Bayesian fusion procedure. The goal is to combine several quickly estimated segmentation maps in order to achieve a more reliable and accurate segmentation result. This fusion is achieved in the penalized maximum PRI sense which has a perceptual meaning. This fusion framework remains simple to implement, perfectible, by increasing the number of segmentation to be fused, and general enough to be applied to various digital image and computer vision applications (e.g., hyperspectral imagery, motion detection, 3-D segmentation, 3-D reconstruction). We think, that this easily parallelizable Bayesian fusion model (thus well suited for the next generation massively parallel computers), appears as an interesting alternative to complex, computationally demanding, segmentation models existing in the literature for the difficult image segmentation problem.

Acknowledgment

The author would like to thank the anonymous reviewers for their many valuable comments and suggestions that helped to improve both the technical content and the presentation quality of this paper.

References

[1] S. Banks, *Signal Processing, Image Processing and Pattern Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1990.

[2] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Stat. Soc.*, vol. B-48, pp. 259–302, 1986.

[3] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[4] Q. Luo and T. Khoshgoftaar, "Unsupervised multiscale color image segmentation based on mdl principle," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2755–2761, Sep. 2006.

[5] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[6] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vision*, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[7] P. Arbelaez and L. Cohen, "A metric approach to vector-valued image segmentation," *Int. J. Comput. Vision, Special Issue Geometrical, Variational, Level Sets Methods Comput. Vision*, vol. 69, no. 1, pp. 119–126, Aug. 2006.

[8] P. Arbelaez, "Boundary extraction in natural images using ultrametric contour maps," in *Proc. 5th IEEE Workshop Perceptual Organization Comput. Vision (POCV)*, New York, Jun. 2006, pp. 182–189.

[9] S. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 9, pp. 884–900, Sep. 1996.

[10] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Sonar image segmentation using a hierarchical MRF model," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1216–1231, Jul. 2000.
[11] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Three-class Markovian segmentation of high resolution sonar images," *Comput. Vision Image Understanding*, vol. 76, no. 3, pp. 191–204, Dec. 1999.
[12] F. Destrempes, J.-F. Angers, and M. Mignotte, "Fusion of hidden Markov Random Field models and its Bayesian estimation," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2920–2935, Oct. 2006.
[13] Z. Kato, T. C. Pong, and G. Q. Song, "Unsupervised segmentation of color textured images using a multi-layer MRF model," in *Proc. Int. Conf. Image Process. (ICIP)*, Barcelona, Spain, Sep. 2003, pp. 961–964.
[14] M. Mignotte, "Segmentation by fusion of histogram-based k-means clusters in different color spaces," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 780–787, May 2008.
[15] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 929–944, Jun. 2007.
[16] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "A measure for objective evaluation of image segmentation algorithms," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recog. (CVPR), Workshop Empirical Evaluation Methods Comput. Vision*, San Diego, CA, Jun. 2005, vol. 3, pp. 34–41.
[17] A. L. N. Fred and A. K. Jain, "Combining multiple clusterings using evidence accumulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 835–850, Jun. 2005.
[18] S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld, "Detection and localization of people in video images using adaptive fusion of color and edge information," in *Proc. Int. Conf. Pattern Recog. (ICPR)*, Barcelona, Spain, Sep. 2000, pp. 4627–4631.
[19] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.
[20] A. Kushki, P. Androutsos, K. Plataniotis, and A. Venetsanopoulos, "Retrieval of images from artistic repositories using a decision fusion framework," *IEEE Trans. Image Process.*, vol. 13, no. 3, pp. 277–292, Mar. 2004.
[21] S. Reed, I. R. Ruiz, C. Capus, and Y. Petillot, "The fusion of large scale classified side-scan sonar image mosaics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 7, pp. 2049–2060, Jul. 2006.
[22] A. J. Sharkey, *Combining Artificial Neural Nets Ensemble and Modular Multi-Net Systems*. New York: Springer-Verlag, 1999.
[23] T. Dietterich, "Ensemble methods in machine learning," in *Proc. 1st Int. Workshop Multiple Classifier Syst., LNCS, Multiple Classifier Syst.*, L. N. I. C. Science, Ed., 2000, vol. 1857, pp. 1–15.
[24] P.-M. Jodoin, M. Mignotte, and C. Rosenberger, "Segmentation framework based on label field fusion," *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 2535–2550, Oct. 2007.
[25] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.
[26] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int. Conf. Comput. Vision (ICCV)*, Vancouver, BC, Canada, Jul. 2001, vol. 2, pp. 416–423.
[27] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.
[28] F. Destrempes, M. Mignotte, and J.-F. Angers, "A stochastic method for Bayesian estimation of hidden Markov models with application to a color model," *IEEE Trans. Image Process.*, vol. 14, no. 8, pp. 1096–1108, Aug. 2005.
[29] M. Mignotte, "Nonparametric multiscale energy-based model and its application in some imagery problems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 184–197, Feb. 2004.
[30] P. Strenski and S. Kirkpatrick, "Analysis of finite length annealing schedules," *Algorithmica*, vol. 6, pp. 346–366, 1991.
[31] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–136, Mar. 1982.
[32] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *Proc. Euro. Conf. Comput. Vision (ECCV), LNCS 2350*, Copenhaguen, Denmark, Jun. 2002, pp. 661–675.
[33] S. B. WesolKowski, "Color image edge detection and segmentation: A comparison of the vector angle and the euclidean distance color similarity measures," M.S. thesis, Syst. Des. Eng. Dept., Univ. Waterloo, Waterloo, ON, Canada, 1999.
[34] E. Maggio and A. Cavallaro, "Multi-part target representation for color tracking," in *Proc. Int. Conf. Image Process. (ICIP)*, Genova, Italy, Sep. 2005, pp. 729–732.
[35] Z. Kato, "A Markov Random Field image segmentation model for color textured images," *Image Vision Comput.*, vol. 24, no. 10, pp. 1103–1114, Oct. 2006.
[36] J.-P. Braquelaire and L. Brun, "Comparison and optimization of methods of color image quantization," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 1048–1052, Jul. 1997.
[37] J. C. Terrillon, M. Shirazi, H. Fukamachi, and S. Akamatsu, "Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recog. (AFGR)*, Grenoble, France, Mar. 2000, pp. 54–61.
[38] J. B. Martinkauppi, M. N. Soriano, and M. H. Laaksonen, "Behavior of skin color under varying illumination seen by different cameras at different color spaces," in *Proc. SPIE, Mach. Vision Appl. Ind. Inspection IX*, San Jose, CA, Jan. 2001, pp. 102–113.
[39] J.-P. Braquelaire and L. Brun, "Comparison and optimization of methods of color image quantization," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 1048–1952, Jul. 1997.
[40] A. Y. Yang, J. Wright, S. Sastry, and Y. Ma, "Unsupervised segmentation of natural images via lossy data compression," *Comput. Vision Image Understanding*, vol. 110, no. 2, pp. 212–225, May 2008.
[41] A. Y. Yang, J. Wright, S. Sastry, and Y. Ma, "Unsupervised segmentation of natural images via lossy data compression," EECS Dept., Univ. California, Berkeley, Tech. Rep. UCB/EECS-2006-195, Dec. 2006. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-195.html
[42] Y. Ma, H. Derksen, W. Hong, and J. Wright, "Segmentation of multivariate mixed data via lossy coding and compression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1546–1562, Sep. 2007.
[43] D. Martin, C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color and texture cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 530–549, May 2004.
[44] D. E. Ilea and P. F. Whelan, "CTEX- an adaptive unsupervised segmentation algorithm on color-texture coherence," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1926–1939, Oct. 2008.
[45] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 8, pp. 800–810, Aug. 2001.
[46] D. Martin, "An empirical approach to grouping and segmentation," Ph.D. dissertation, Elect. Eng. Comput. Sci. Dept., Univ. California, Berkeley, Aug. 2002.
[47] D. Martin and C. Fowlkes, , "The Berkeley Segmentation Database and Benchmark," Univ. California, Berkeley, 2010. [Online]. Available: http//www.cs.berkeley.edu/projects/vision/grouping/segbench/
[48] M. Maire, P. Arbelaez, C. Fowlkes, and J. Malik, "Using contours to detect and localize junctions in natural images," in *Proc. IEEE Conf. Comput. Vision Pattern Recog. (CVPR)*, Anchorage, AK, Jun. 2008, pp. 1–8.
[49] P.-M. Jodoin and M. Mignotte, "Markovian segmentation and parameter estimation on graphics hardware," *J. Electron. Imag.*, vol. 15, no. 3, pp. 033 015-1–033 015-15, Jul.–Sep. 2006.

Max Mignotte received the DEA (Postgraduate degree) in digital signal, image and speech processing from the INPG University, Grenoble, France, in 1993 and the Ph.D. degree in electronics and computer engineering from the University of Bretagne Occidentale (UBO) and the digital signal laboratory (GTS) of the French Naval Academy, France, in 1998.

He was an INRIA Post-Doctoral Fellow with University of Montreal (DIRO), Montreal, QC, Canada, from 1998 to 1999. He is currently with the Computer Vision & Geometric Modeling Lab, DIRO, University of Montreal, as an Associate Professor (Professeur agrégé). He is also a member of Laboratoire de recherche en imagerie et orthopedie, Centre de recherche du CHUM, Hopital Notre-Dame (LIO) and Researcher at CHUM. His current research interests include statistical methods, Bayesian inference and hierarchical models for high-dimensional inverse problems such as segmentation, parameters estimation, fusion, shape recognition, deconvolution, 3-D reconstruction, and restoration problems.